Université de Parakou (UP), Bénin

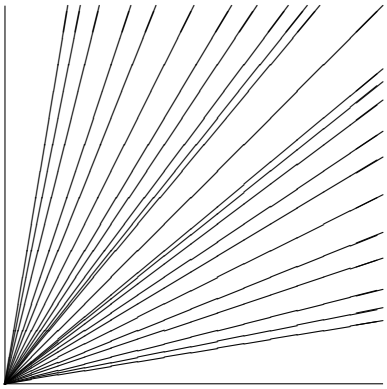
Institut de Mathématiques et de Sciences Physiques (IMSP-BENIN)

Ecole Normale Supérieure de Natitingou (ENS)

Titre :

COURS DE MESURE ET INTEGRATION

Pour :
Professeur Marcos ABOUBACAR
(abmarcos@yahoo.fr)
Soumis par :
Nonvikan Karl-Augustt Alahassa
(karl.augustt.corporate@gmail.com)



Année académique : 2011 - 2012

©ENS/FAST/UP 2012

Table des matières

1 Fonctions étagées **4**
- 1.1 Fonctions étagées . . . 4
- 1.2 Mesures images et lois de probabilité . . . 5

2 Integration **7**
- 2.1 Intégration des fonctions mesurables positives . . . 7
- 2.2 Intégration d'une fonction mesurable quelconque . . . 8
- 2.3 Théorèmes de convergence . . . 10
- 2.4 Intégration par rapport à une mesure image . . . 12

3 Intégration sur les espaces produits **15**
- 3.1 Mesures de probabilité produit . . . 15

4 Dépendance et indépendance **19**
- 4.1 Indépendance d'évènements . . . 19
- 4.2 Indépendance de classes . . . 21
- 4.3 Indépendance de variables aléatoires . . . 23
- 4.4 Dépendance de variables aléatoires . . . 29
- 4.5 Convolution . . . 32

5 Lois de probabilité absolument continues - Lois de probabilité conditionnelle **35**
- 5.1 Lois de probabilité absolument continues . . . 35
- 5.2 Lois de probabilité conditionnelle – Espérance mathématique conditionnelle . . . 37
- 5.3 Règles de Calcul sur les espérances conditionnelles . . . 39

6 Fonctions génératrices – fonctions caractéristiques **44**
- 6.1 Fonction génératrice . . . 44
- 6.2 Applications . . . 48
- 6.3 Fonctions caractéristiques . . . 52

7 Lois de probabilité de fonctions de variables aléatoires **55**
- 7.1 Cas à une dimension . . . 55
- 7.2 Cas à deux dimensions . . . 57

8 Convergences stochastiques **60**
- 8.1 Convergence en loi ou convergence étroite . . . 60
- 8.2 Convergence en probabilité . . . 62

©ENS/FAST/UP 2012

8.3 Convergence en moyenne d'ordre r 64
8.4 Convergence presque sûre 64
8.5 Comparaison des divers types de convergence 66
8.6 Loi des grands nombres 67
8.7 Théorème *"central limite"* 68

 ©ENS/FAST/UP 2012

CHAPITRE PREMIER

FONCTIONS ÉTAGÉES

1.1 Fonctions étagées

Définition 1.1. *Soit f une fonction de $\mathcal{E}$ dans $\mathbb{R}$. La fonction f est dite fonction étagée si l'ensemble $f(\mathcal{E})$ est fini, c'est-à-dire $f(\mathcal{E}) = \{y_1, y_2, ..., y_n\}$ avec $y_i \in \mathbb{R}$, $1 \leq i \leq n$. Si en plus f est mesurable de $(\mathcal{E}, \tau)$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, où $(\mathcal{E}, \tau)$ est un espace mesurable, on dit que f est une fonction étagée mesurable.*

Exemple

Soit $A \subset \mathcal{E}$, et

$$\begin{array}{rccl} 1_A & : \mathcal{E} & \longrightarrow & \mathbb{R} \\ & x & \longmapsto & \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases} \end{array}$$

1_A une fonction étagée. C'est la fonction indicatrice de A. La fonction 1_A est mesurable si et seulement si A est mesurable.

Proposition 1.1. *Soit f une fonction étagée de $\mathcal{E}$ dans $\mathbb{R}$ prenant les valeurs y_1, y_2,..., y_n, alors on a la décomposition canonique : $f = \sum_{i=1}^{n} y_i 1_{A_i}$, avec $A_i = f^{-1}(\{y_i\})$.*

Démonstration. En effet les $(A_i)_{1 \leq i \leq n}$, forment une partition de $\mathcal{E}$, c'est-à-dire $\mathcal{E} = \bigcup_{i=1}^{n} A_i$, et $A_i \cap A_j = 0 \quad \text{si } i \neq j$. Ainsi, $x \in \mathcal{E} \iff \exists k \text{ unique}, 1 \leq k \leq n \quad \text{tel que } x \in A_k$. Or $x \in A_k \iff \sum_{i=1}^{n} y_i 1_{A_i}(x) = y_k 1_{A_k} = f(x)$. Donc $f(x) = \sum_{i=1}^{n} y_i 1_{A_i}(x)$, $\forall x \in \mathcal{E}$.

□

Corollaire 1. *Soit $(\mathcal{E}, \tau)$ un espace mesurable et f une fontion de $\mathcal{E}$ dans $\mathbb{R}$, étagée et prenant les valeurs y_1, y_2, ..., y_n. f est mesurable si et seulement si les ensembles $A_i = f^{-1}(\{y_i\})$, $i \in \{1; 2; ...; n\}$, sont mesurables.*

Théorème 1.1. *(Admis)*
Soit $(\mathcal{E}, \tau)$ un espace mesurable et f une fonction de $\mathcal{E}$ dans $\bar{\mathbb{R}}_+$ mesurable. Alors f est limite simple d'une suite croissante de fonctions étagées mesurables positives.

Remarque 1.1. *Dans le théorème précédent , la convergence est uniforme sur $\mathcal{E}$, si f est bornée.*

Corollaire 2. *Soit $(\mathcal{E}, \tau)$ un espace mesurable et f une fonction de $\mathcal{E}$ dans $\bar{\mathbb{R}}$, mesurable. Alors f est limite simple d'une suite de fonctions étagées mesurable.*

©ENS/FAST/UP 2012

Démonstration. On pose $f = f^+ - f^-$, et on applique le théorème pécédent à f^+ et à f^- qui sont des fonctions positives mesurables. □

Remarque 1.2. *Tout ce qui vient d'être établi pour les fonctions mesurables est aussi valable pour les variables aléatoires.*

1.2 Mesures images et lois de probabilité

Les applications mesurables permettent de munir l'espace d'arrivée d'une mesure à partir de la mesure sur l'espace de départ. En particulier une variable aléatoire permet de probabiliser l'espace d'arrivée à partir de la probabilité sur l'espace de départ.

Proposition 1.2. *Soient $(\mathcal{E}, \tau)$ et $(\mathcal{E}', \tau')$ deux espaces mesurables, μ une mesure sur $(\mathcal{E}, \tau)$ et f une application de $(\mathcal{E}, \tau)$ dans $(\mathcal{E}', \tau')$ mesurable. La fonction d'ensembles $\mu \circ f^{-1}$ définie sur τ' par :*

$$\forall B \in \tau', (\mu \circ f^{-1})(B) = \mu(f^{-1}(B))$$

est une mesure sur $(\mathcal{E}', \tau')$. On l'appelle mesure image de μ par f.

Démonstration. $\nu = \mu \circ f^{-1}$ est bien définie sur τ' puisque pour tout B élément de τ', $f^{-1}(B) \in \tau$ et μ est définie sur τ. Par ailleurs, $\nu(\phi) = \mu(f^{-1}(\phi)) = \mu(\phi) = 0$.
Soit $(B_n)_n$ une suite d'éléments de τ' deux à deux disjoints.

$$\nu(\bigcup_n B_n) = \mu[\bigcup_n f^{-1}(B_n)] = \sum_n \mu(f^{-1}(B_n)) = \sum_n \nu(B_n)$$

□

Définition 1.2. *Soit (Ω, τ, P) un espace probabilité et $X :\longrightarrow \mathbb{R}^d$, un vecteur aléatoire. La mesure image $P_X = P \circ X^{-1}$ qui est une probabilité sur $\mathcal{B}(\mathbb{R}^d)$ est appelée loi de probabilité du vecteur aléatoire X (On dit loi de probabilité de la variable aléatoire X si $d = 1$). Ainsi, $\forall B \in \mathcal{B}(\mathbb{R}^d)$, $P_X(B) = P(X^{-1}(B))$.*

Démonstration. Découle de la proposition N° ? et du fait que $P_X(\mathbb{R}^d) = P(\Omega) = 1$. □

Notation

$P_X(B) = P(X \in B)$.

Remarque 1.3. *Soit $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), P')$ un espace probabilisé, on peut lui associer un espace probabilisé (Ω, τ, P) et une variable aléatoire $X : (\Omega, \tau) \longrightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ telle que $P_X = P'$. Il suffit de prendre $\Omega = \mathbb{R}^d$, $\tau = \mathcal{B}(\mathbb{R}^d)$, $P = P'$, et $X = Id_{\mathbb{R}}$ (application identique).*

1. *Deux variables aléatoires X et X' définies sur le même espace peuvent avoir même loi sans être égales, voire avoir même loi (variables équidistribuées), sans être définies sur le même espace probabilisé.*
2. *Dans beaucoup de problèmes une variable aléatoire X n'intervient que par sa loi P_X, l'espace (Ω, τ, P) n'étant pas précisé ; alors, parmi les variables aléatoires équidistribuées ayant la loi de probabilité P_X, ou en choisit une .*

3. *Enfin, rappelons que si (Ω, τ, P) est un espace probabilisé et $H \in \tau$ un événement tel que $P(H) > 0$, on peut définir sur τ une nouvelle mesure de probabilité $P_H = P(./H)$ par :*

$$\forall A \in \tau, P_H(A) = P(A/H) = \frac{P(A \cap H)}{P(H)}$$

Si X est un vecteur aléatoire sur (Ω, τ), la probabilité P_H induit sur X une loi de la manière suivante.

Définition 1.3. *Soient (Ω, τ, P) un espace probabilisé, $H \in \tau$ tel que $P(H) > 0$, X un vecteur aléatoire de (Ω, τ) dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. On appelle loi conditionnelle de X sachant H, la loi $P_{X/H}$ définie par :*

$$\forall B \in \mathcal{B}(\mathbb{R}^d), P_{X/H}(B) = P_H(X^{-1}(B)) = P(X \in B/H)$$

©ENS/FAST/UP 2012

Integration

2.1 Intégration des fonctions mesurables positives

On supposera dans tout ce qui suit que $(\mathcal{E}, \tau, \mu)$ est un espace mesuré.

2.1.1 Intégration de fonctions étagées positives mesurables

Définition 2.1. *Soit f une fonction étagée mesurable positive, prenant les valeurs y_1, y_2, y_n. $f = \sum_{i=1}^{n} y_i 1_{A_i}$, avec $A_i = f^{-1}(\{y_i\})$. Soit B inτ, on définit l'intégrale de f sur B par rapport à μ par :*

$$\int_B f d\mu = \sum_{i=1} n y_i . \mu(A_i \cap B)$$

Proposition 2.1. *(*Admis*)*

1. *Soient f et g deux fonctions étagées mesurables positives, alors :*

$$f \leq g \Longrightarrow \int_A f d\mu \leq \int_A g d\mu, \ \ \forall A \in \tau$$

2. *Soient $(f_n)_n$, $(g_n)_n$ deux suites croissantes de fonctions étagées mesurables positives, alors*

$$\sup_n f_n = \sup_n g_n \Longrightarrow \sup_n \int_A f_n d\mu = \sup_n \int_A g_n d\mu, \ \ \forall A \in \tau.$$

2.1.2 Intégration de fonctions mesurables positives

Définition 2.2. *Soit f une fonction mesurable positive à valeurs dans $\bar{\mathbb{R}}_+$. On définit l'intégrale de f par rapport à μ sur un ensemble de A de τ par :*

$$\int_A f d\mu = \sup \int_A s d\mu$$

où le supremum est pris sur l'ensemble des fonctions étagées mesurables positives telle que $0 \leq s \leq f$. On dit que f est μ-intégrable sur A si $\int_A f d\mu < +\infty$.

Remarque 2.1. *Il est équivalent de définir l'intégrale de f comme suit :*

$$\int_A f d\mu = \lim_{n \to +\infty} \int_A s_n d\mu$$

où $(s_n)_n$ est une suite croissante de fonctions étagées mesurables positive.

©ENS/FAST/UP 2012

2.2 Intégration d'une fonction mesurable quelconque

Définition 2.3. *Soit f une fonction mesurable à valeurs dans $\bar{\mathbb{R}}$. On dit que f est μ-intégrable sur A élément de τ si $\int_A f^+ d\mu$, $\int_A f^- d\mu$ sont toutes deux finies. Le réel*

$$\int_A f^+ d\mu - \int_A f^- d\mu = \int_A f d\mu$$

est appelé intégrale de f par rapport à μ sur A . L'ensemble des fonctions μ-intégrables sur $\mathcal{E}$ est noté $\mathcal{L}^1_\mu(\mathcal{E})$.

Remarque 2.2.

$$f \in \mathcal{L}^1_\mu(\mathcal{E}) \Longleftrightarrow |f| \in \mathcal{L}^1_\mu(\mathcal{E}), \text{ puisque } |f| = f^+ + f^-$$

Exemple : Mesure de Dirac et de dénombrement

1. Soit $(\mathcal{E}, \tau, \delta_a)$ un espace mesuré où δ_a est la mesure de Dirac au point $a \in \mathcal{E}$ et soit f une fonction positive mesurable sur $\mathcal{E}$. Alors,

$$\int_{\mathcal{E}} f d\delta_a = f(a) \text{ (à montrer en exercice)}$$

2. Soit μ_d la mesure de dénombrement sur $\mathcal{P}(\mathbb{N})$ et f une fonction sur $\mathbb{N}$ mesurable et positive. On a :

$$\int_{\mathbb{N}} f d\mu_d = \sum_{p=1}^{\infty} \text{ (à montrer en exercice)}$$

Proposition 2.2. *Soient f et g deux fonctions mesurables sur $(\mathcal{E}, \tau)$ et à valeurs dans $\bar{\mathbb{R}}$. On suppose que $f \geq 0$, $g \geq 0$, ou que f et g sont quelconques mais μ-intégrables sur $\mathcal{E}$. Alors on a :*

1. $\int_{\mathcal{E}} (f+g) d\mu = \int_{\mathcal{E}} f d\mu + \int_{\mathcal{E}} g d\mu$
2. *Pour tout réel λ, on a :* $\int_{\mathcal{E}} \lambda f d\mu = \lambda \int_{\mathcal{E}} f d\mu$
3. *Si $f \leq g$, alors* $\int_{\mathcal{E}} f d\mu \leq \int_{\mathcal{E}} g d\mu$
4. *Si $A \subset \mathcal{E}$, alors,* $\int_A f d\mu = \int_{\mathcal{E}} 1_A f d\mu$
5. *Si $A \subset B \subset \mathcal{E}$ et si $f \geq 0$, alors* $\int_A f d\mu \leq \int_B f d\mu$
6. *Si $f = 0$ sur $\mathcal{E}$ alors* $\int_{\mathcal{E}} f d\mu = 0$, *même si* $\mu(\mathcal{E}) = +\infty$
7. *Si $\mu(\mathcal{E}) = 0$, alors* $\int_{\mathcal{E}} f d\mu = 0$, *même si* $f(x) = \infty, \forall x \in \mathcal{E}$
8. *Si f est μ-intégrable, alors* $|\int_{\mathcal{E}} f d\mu| \leq \int_{\mathcal{E}} |f| d\mu$.

Démonstration. De n°1 à n°7, résultent de l'application des définitions d'intégrales,

$$n8) \ |\int_{\mathcal{E}} f d\mu| = |\int_{\mathcal{E}} f^+ d\mu - \int_{\mathcal{E}} f^- d\mu| \leq \int_{\mathcal{E}} f^+ d\mu + \int_{\mathcal{E}} f^- d\mu = \int_{\mathcal{E}} |f| d\mu$$

□

Proposition 2.3. *Soit f une fonction étagée positive mesurable sur $(\mathcal{E}, \tau)$. Pour tout A élément de τ, l'application $\nu : \tau \longrightarrow \bar{\mathbb{R}}_+$ définit par :*

$$\nu(A) = \int_A f d\mu$$

est une mesure sur τ. On dit que ν est de densité f par rapport à μ.

Démonstration.

ν est évidemment bien définie car $f \geq 0$ et $\nu(\phi) = 0$.

Posons $f(\mathcal{E}) = \{y_1, y_2, ..., y_n\}$. Alors,

$$f = \sum_{i=1}^{n} y_i 1_{A_i} \text{ avec } A_i = f^{-1}(\{y_i\})$$

$$(A) = \int_A f d\mu = \sum_{i=1}^{n} y_i \mu(A_i \cap A)$$

Soit $(\mathcal{E}_j)_j$ une famille dénombrable d'éléments de τ deux à deux disjoints.

$$\nu(\bigcup_{j=1}^{\infty} \mathcal{E}_j) = \sum_{i=1}^{n} y_i \mu(A_i \cap (\bigcup_{j=1}^{\infty} \mathcal{E}_j)) = \sum_{i=1}^{n} y_i \mu[\bigcup_{j=1}^{\infty} (A_i \cap \mathcal{E}_j)] = \sum_{i=1}^{n} y_i \sum_{j=1} \infty \mu(A_i \cap \mathcal{E}_j)$$

Car la famille $(A_i \cap \mathcal{E}_j)_j$ est une famille dénombrable d'éléments de τ deux à deux disjoints. Ainsi,

$$\begin{aligned} \nu(\bigcup_{j=1}^{\infty} \mathcal{E}_j) &= \sum_{i=1}^{n} y_i \sum_{j=1}^{\infty} \mu(A_i \cap \mathcal{E}_j) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{\infty} y_i \mu(A_i \cap \mathcal{E}_j) \\ &= \sum_{j=1}^{\infty} \nu(\mathcal{E}_j) \end{aligned}$$

D'où ν est une mesure. □

Remarque 2.3. *La proposition* 2.3. *est encore vraie si f est une fonction mesurable positive à valeurs dans* $\bar{\mathcal{E}}_+$.

Définition 2.4. *$(\mathcal{E}, \tau, \mu)$ étant un espace mesuré, une propriété $\mathcal{P}$ dont la valeur de vérité dépend de x élément de $\mathcal{E}$ est dite vraie μ-presque partout $(\mu - pp)$, s'il existe $A \in \tau$, $\mu(A) = 0$, et tel que $\mathcal{P}$ est vraie pour tout $x \in A^c$.*

Proposition 2.4.

Soit f une fonction positive mesurable sur $(\mathcal{E}, \tau)$ à valeurs dans $\bar{\mathbb{R}}_+$.

$$\int_{\mathcal{E}} f d\mu = 0 \Longleftrightarrow f = 0, \mu - pp$$

Démonstration. 1. Supposons $f = 0$ sur $\mathcal{E}/A$, avec $\mu(A) = 0$.

$$\int_{\mathcal{E}} f d\mu = \int_{\mathcal{E}/A} f d\mu + \int_A f d\mu = 0$$

©ENS/FAST/UP 2012

2. Supposons $\int_{\mathcal{E}} f d\mu = 0$.
   Posons $B_n = \{x \in \mathcal{E}, f(x) \geq \frac{1}{n}\}$

$$0 = \int_{\mathcal{E}} f d\mu \geq \int_{B_n} f d\mu \geq \frac{1}{n}\mu(B_n)$$

   Ce qui implique que $\mu(B_n) = 0$. Or $\{x, f(x) > 0\} = \bigcup_{n\geq 1} B$, donc $\mu\{x/f(x) > 0\} \leq \sum_{n\geq 1} \mu(B_n) = 0$, donc $f = 0$, $\mu - pp$.

□

Proposition 2.5.

Soient f et g deux fonctions positives mesurables sur $(\mathcal{E}, \tau)$, à valeurs dans $\bar{\mathbb{R}}_+$.

$$f = g, \mu - pp \Longrightarrow \int_{\mathcal{E}} f d\mu = \int_{\mathcal{E}} g d\mu$$

.

Démonstration. $f = g\,\mu - pp$ implique qu'il existe $A \in \tau$, $\mu(A) = 0$, et $f = g$ sur $\mathcal{E}/A$.

$$\begin{aligned}
\int_{\mathcal{E}} f d\mu &= \int_{\mathcal{E}/A} f d\mu + \int_A f d\mu \\
&= \int_{\mathcal{E}/A} f d\mu \\
&= \int_{\mathcal{E}/A} g d\mu + \int_A g d\mu \\
&= \int_{\mathcal{E}} g d\mu
\end{aligned}$$

A présent nous abordons les grands théorèmes d'intervension limite-intégrale. Nous les énoncerons sans donner leurs démonstrations. □

2.3 Théorèmes de convergence

Théorème 2.1. *(De convergence monotone ou de Beppolevi)*
Soit $(f_n)_n$ une suite croissante de fonctions mesurables sur $(\mathcal{E}, \tau)$ à valeurs dans $\bar{\mathbb{R}}_+$ convergeant simplement vers f sur $\mathcal{E}$. Alors on a :

$$\int_{\mathcal{E}} (\lim_{n\to+\infty} f_n) d\mu = \lim_{n\to+\infty} \int_{\mathcal{E}} f_n d\mu$$

Autrement dit,

$$\int_{\mathcal{E}} f d\mu = \lim_{n\to+\infty} \int_{\mathcal{E}} f_n d\mu$$

cette égalité ayant lieu dans $[0; +\infty]$.

Corollaire 3. *Soit $(f_n)_n$ une suite de fonctions mesurables sur $(\mathcal{E}, \tau)$, à valeurs dans $\bar{\mathbb{R}}_+$. Alors :*

$$\int_{\mathcal{E}} (\sum_{n\geq 1} f_n) d\mu = \sum_{n\geq 1} \int_{\mathcal{E}} f_n d\mu$$

.

Démonstration. Considérons la suite de terme général $g_k = \sum_{n=1} kf_n$. $(g_k)_k$ est une suite croissante de fonctions positives mesurables sur $(\mathcal{E}, \tau)$ qui converge vers $\sum_{n\geq 1} f_n$. On applique alors le théorème de Beppo-Levi à la suite $(g_k)_k$, et on a le résultat.

□

Proposition 2.6. *$(\mathcal{E}, \tau, \mu)$ étant un espace mesuré et ν une mesure de densité f par rapport à μ. Soit g une fonction mesurable sur $(\mathcal{E}, \tau)$ à valeurs dans $\bar{\mathbb{R}}_+$, alors on a :*

$$\int_{\mathcal{E}} g d\nu = \int_{\mathcal{E}} gf d\nu$$

Démonstration. Si $g = 1_A$, $A \in \tau$,

$$\int_{\mathcal{E}} g d\nu = \int_{\mathcal{E}} 1_A d\nu = \nu(A) = \int_A f d\mu = \int_{\mathcal{E}} 1_A f d\mu$$

. Si $g = \sum_{i=1}^n y_i 1_{A_i}$, on a :

$$\begin{aligned}
\int_{\mathcal{E}} g d\mu &= \sum_{i=1}^n y_i \nu(A_i) \\
&= \sum_{i=1}^n y_i \int_{A_i} f d\mu \\
&= \sum_{i=1}^n y_i \int_{\mathcal{E}} 1_{A_i} f d\mu \\
&= \int_{\mathcal{E}} \sum_{i=1}^n y_i 1_{A_i} f d\mu \\
&= \int_{\mathcal{E}} gf d\mu
\end{aligned}$$

A présent supposons que g est une fonction mesurable quelconque à valeurs dans $\bar{\mathbb{R}}_+$. D'après le théorème ? du chapître 2, il existe une suite croissante de fonctions étagées positives convergeant vers g. Notons $(g_n)_n$ cette suite. La fonction f étant positive la suite $(g_n f)_n$ est une suite croissante de fonctions mesurables. Montrons que $(g_n f)_n$ converge vers (gf).
Si le couple (g, f), ne prend pas les valeurs $(0, +\infty)$, ou $(+\infty, 0)$, alors la suite $(g_n f)_n$ converge vers (gf).
Voyons maintenant le cas où il existe un $x \in \mathcal{E}$ tel que $g(x) = 0$ et $f(x) = +\infty$, ou $g(x) = +\infty$ et $f(x) = 0$.
Si $g(x) = 0$ et $f(x) = +\infty$, alors la suite $(g_n)_n$ étant croissante $g_n = 0$, $\forall n$, et alors $g_n(x)f(x) = 0$, d'où $g_n(x)f(x)$ converge vers $g(x)f(x)$.

Si $g(x) = +\infty$ et $f(x) = 0$, alors $g_n(x)f(x) = 0$, $\forall n$, et alors $g_n(x)f(x)$ converge encore vers $g(x)f(x)$. Les suites $(g_n)_n$ et $(g_n f)_n$ vérifiant les hypotèses du théorème de Beppo-Lévi. On a :

$$\lim_{n\to+\infty} \int_{\mathcal{E}} g_n d\nu = \int_{\mathcal{E}} g d\nu$$

et

$$\lim_{n\to+\infty} \int_{\mathcal{E}} g_n d\mu = \int_{\mathcal{E}} g d\mu$$

Les g_n étant des fonctions étagées, $\int_{\mathcal{E}} g_n d\nu = \int_{\mathcal{E}} g d\mu$, et donc par unicité de la limite, on a : $\int_{\mathcal{E}} g d\mu = \int_{\mathcal{E}} g f d\mu$ □

Théorème 2.2. *(Lemme de Fatou)*
Soit $(f_n)_n$ une suite de fonctions mesurables sur $(\mathcal{E}, \tau)$ à valeurs dans $\bar{\mathbb{R}}_+$. Alors on a :

$$\int_{\mathcal{E}} (\underline{\lim}_n f_n) d\mu \leq \underline{\lim}_n \int_{\mathcal{E}} f_n d\mu$$

Théorème 2.3. *(Convergence dominée de Lebesgue)*
Soit $(f_n)_n$ une suite de fonctions mesurables sur $(\mathcal{E}, \tau)$ à valeurs dans $\bar{\mathbb{R}}$. On suppose que :

1. *f_n converge vers f presque partout.*
2. *Il eiste une fonction mesurable positive g μ-intégrable telle que $|f_n| \leq g$ $\forall$.*

Alors la suite $(f_n)_n$ et la fonction f sont intégrables et on a :

$$\lim_{n\to+\infty} \int_{\mathcal{E}} f_n d\mu = \int_{\mathcal{E}} \lim_{n\to+\infty} f_n d\mu = \int_{\mathcal{E}} f d\mu$$

2.4 Intégration par rapport à une mesure image

Soient $(\mathcal{E}, \tau)$, $(\mathcal{E}', \tau')$ deux espaces mesurables, et soit ϕ une application mesurable de $(\mathcal{E}, \tau)$ dans $(\mathcal{E}', \tau')$.

Théorème 2.4. *(Théorème de transfert)*

Soit μ une mesure sur $(\mathcal{E}, \tau)$ et $\nu = \mu \circ \phi^{-1}$ sa mesure image par ϕ sur $(\mathcal{E}', \tau')$. Alors pour toute fonction $h : \mathcal{E}' \longrightarrow \bar{\mathbb{R}}_+$ mesurables, on a :

$$\int_E (h \circ \phi) d\mu = \int_{\mathcal{E}} h d\nu$$

Démonstration. $h \circ \phi$ est évidemment mesurable de $(\mathcal{E}, \tau)$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Si $h = 1_B$ avec $B \in \tau'$, alors $A = \phi^{-1}(B) \in \tau$, et on a :

$$\int_E 1_B d\nu = \nu(B) = \mu(\phi^{-1}(B)) = \mu(A) = \int_{\mathcal{E}} 1_A d\mu$$

Montrons que $1_A = 1_B \circ \phi$. Soit $x \in \mathcal{E}$,

$$1_A(x) = \begin{cases} 1 & \text{si } x \in A = \phi^{-1}(B) \\ 0 & \text{si } x \in x \notin A \end{cases} = \begin{cases} 1 & \text{si } \phi(x) \in B \\ 0 & \text{si } \phi(x) \notin B \end{cases}$$

$$1_B \circ \phi(x) = 1_B(\phi(x)) = \begin{cases} 1 & \text{si } \phi(x) \in B \\ 0 & \text{si } \phi(x) \notin B \end{cases}$$

Il en résulte que $1_A = 1_B \circ \phi$. Si h est une fonction étagée, $h = \sum_{i=1}^n y_i 1_{B_i}$, alors $h \circ \phi = \sum_{i=1}^n y_i (1_{B_i} \circ \phi)$.

$$\begin{aligned}\int_{\mathcal{E}}(h\circ\phi)d\mu &= \sum_{i=1}^{n} y_i \int_{\mathcal{E}}(1_{B_i}\circ\phi)d\mu \\ &= \sum_{i=1}^{n} y_i \int_{\mathcal{E}'} 1_{B_i} d\nu \\ &= \int_{\mathcal{E}'}(\sum_{i=1}^{n} y_i 1_{B_i})d\nu \\ &= \int_{\mathcal{E}'} h d\nu\end{aligned}$$

Si h est une fonction mesurable quelconque, alors il existe une suite $(h_n)_n$ de fonctions étégées positives mesurables (la suite étant croissante) convergeant vers h. Il en résulte que la suite $(h_n \circ \phi)_n$ est une suite croissante de fonctions étagées positives mesurables convergeant vers $h \circ \phi$.
Par ailleurs,

$$\int_{\mathcal{E}}(h_n\circ\phi)d\mu = \int_{\mathcal{E}'} h_n d\nu$$

En appliquant le théorème de Beppo-Levi aux suites $(h_n \circ \phi)_n$ et $(h_n)_n$, on a :

$$\begin{aligned}\int_{\mathcal{E}}(h\circ\phi)d\mu &= \lim_{n\to+\infty}\int_{\mathcal{E}}(h_n\circ\phi)d\mu \\ &= \lim_{n\to+\infty}\int_{\mathcal{E}} h_n d\nu \\ &= \int_{\mathcal{E}} h d\nu\end{aligned}$$

Ci-dessous nous donnons une importante généralisation du théorème de transfert utile aux calculs d'espérences. □

Théorème 2.5. *(De transfert)*
Soient $(\mathcal{E},\tau)$ et $(\mathcal{E}',\tau')$ deux espaces mesurables, ϕ une application mesurable de $(\mathcal{E},\tau)$ dans $(\mathcal{E}',\tau')$, μ une mesure sur $(\mathcal{E},\tau)$ et $\nu = \mu\circ\phi^{-1}$ sa mesure image sur $(\mathcal{E}',\tau')$. Soit une fonction $h : \mathcal{E}' \longrightarrow \mathbb{R}$ ou $\bar{\mathbb{R}}$ mesurable, alors h est ν-intégrable si et seulement si $h\circ\phi$ est μ-intégrable, et on a :

$$\int_{\mathcal{E}}(h\circ\phi)d\mu = \int_{\mathcal{E}'} h d\nu$$

Démonstration. $h\circ\phi$ est μ-intégrable si et seulement si $(h\circ\phi)^+$ et $(h\circ\phi)^-$ sont μ-intégrables. Or $(h\circ\phi)^+ = h^+\circ\phi$ et $(h\circ\phi)^- = h^-\circ\phi$. Ainsi,

$$(h\circ\phi)^+ \ \mu-\text{intégrable} \Longleftrightarrow h^+\circ\phi \ \mu-\text{intégrable}$$

$$(h\circ\phi)^- \ \mu-\text{intégrable} \Longleftrightarrow h^-\circ\phi \ \mu-\text{intégrable}$$

En appliquant le théorème de transfert, il en résulte que :

$$h^+ \circ \phi \ \mu - \text{intégrable} \Longleftrightarrow h^+ \ \nu - \text{intégrable (1)}$$
$$h^- \circ \phi \ \mu - \text{intégrable} \Longleftrightarrow h^- \ \nu - \text{intégrable (2)}$$

$$\text{(1) et (2)} \Longleftrightarrow h \ \nu - \text{intégrable}$$

Ainsi $h \circ \phi\mu - \text{intégrable} \Longleftrightarrow h \ \nu - \text{intégrable}$

$$\begin{aligned}\int_{\mathcal{E}}(h\circ\phi)d\mu &= \int_{\mathcal{E}}(h^+\circ\phi)d\mu - \int_{\mathcal{E}}(h^-\circ\phi)d\mu \\ &= \int_{\mathcal{E}}(h^+d\nu - \int_{\mathcal{E}'}h^-d\nu \\ &= \int_{\mathcal{E}'}hd\nu\end{aligned}$$

□

Corollaire 4. *(Calculs d'espérances et de moments)*
Soit (Ω, τ, P) un espace probabilisé, et $X : \Omega \longrightarrow \mathbb{R}^d$ est un vecteur aléatoire et $h : \mathbb{R}^d \longrightarrow \mathbb{R}$ une application borélienne. Si la variable aléatoire $Y = h(x)$ est P-intégrable, on a :

$$Eh(x) = \int_\Omega h(x)dP = \int_{\mathbb{R}^d} h(x)dP_X(x) = \int_{\mathbb{R}} ydP_Y(y)$$

Démonstration. IL suffit d'appliquer le théorème précédent avec $\mathcal{E} = \Omega$, $\mathcal{E}' = \mathbb{R}^d$, $\tau' = \mathcal{B}(\mathbb{R}^d)$, et $\phi = X$.

□

Remarque 2.4. *Le cas où X est une variable aléatoire et $h(x) = x^k$ $(k \in \mathbb{N}^*)$ donne les moments d'ordre k, $E(X^k)$ de X et le cas particulier $k = 1$ donne l'espérance mathématique $E(X)$ de X.*

Corollaire 5. *(Espérance d'une variable aléatoire discrète)*

Soit (Ω, τ, P) un espace probabilisé et X une variable aléatoire discrète sur (Ω, τ) à valeurs dans $\mathbb{R}$. Soit h une application quelconque de $\mathbb{R}$ dans $\mathbb{R}$, alors $h \circ X$ est une variable aléatoire discrète et si $h \circ X$ est P-intégrable, alors : $Eh(X) = \sum h(x)P(X = x)$.

Démonstration. $X(\Omega)$ est une partie au plus dénombrable de $\mathbb{R}$ et $P_X = \sum_{x\in X(\Omega)} P(X = x)\omega_x$.

X étant mesurable de (Ω, τ) dans $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$, $h \circ X$ est mesurable puisque h est automatiquement mesurable de $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ dans $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$. On applique donc le corollaire précédent avec $\mathcal{B}(\mathbb{R}) = \mathcal{P}(\mathbb{R})$.

□

Remarque 2.5. *X étant un vecteur aléatoire, si la loi P_X a une densité f par rapport à la mesure de Lebesgue λ_d sur $\mathbb{R}^d$, l'expression de $Eh(X)$ dans le corollaire n° ? devient :*

$$Eh(x) = \int_{\mathbb{R}^d} hfd\lambda_d$$

 ©ENS/FAST/UP 2012

Intégration sur les espaces produits

3.1 Mesures de probabilité produit

Définition 3.1. *Soit Ω_1, Ω_2 deux ensembles et $\mathcal{E}$ une partie de $\Omega_1 \times \Omega_2$. On appelle section de $\mathcal{E}$ par $\omega_1 \in \Omega_1$ le sous-ensemble :*

$$E\omega_1 = \{\omega_2 \in \Omega_2, (\omega_1, \omega_2) \in \mathcal{E}\}$$

. De même, on appelle section de $\mathcal{E}$ par $\omega_2 \in \Omega_2$, le sous-ensemble :

$$E^{\omega_2} = \{\omega_1 \in \Omega_1, (\omega_1, \omega_2) \in \mathcal{E}\}$$

Proposition 3.1. *Soit $(\Omega_1 \times \Omega_2, \tau_1 \otimes \tau_2)$ un espace mesurable. Soit E un élément de $(\tau_1 \otimes \tau_2)$, alors $\forall \omega_1 \in \Omega_1$, $E_{\omega_1} \in \tau_2$; $\forall \omega_2 \in \Omega_2$, $E_{\omega_2} \in \tau_1$.*

Démonstration. Il suffit de faire la preuve pour la section par ω_1.

Soit $\omega_1 \in \Omega_1$, posons $\tau = \{E \in \tau_1 \otimes \tau_2, E_{\omega_1} \in \tau_2\}$. Montrons que τ est une tribu et que $\tau = \tau_1 \otimes \tau_2$.

Soit $A \in \tau_1$, soit $B \in \tau_2$.

$$(A \times B)_{\omega_1} = \begin{cases} B & \text{si } \omega_1 \in A \\ \phi & \text{si } \omega_2 \notin A \end{cases}$$

Donc $A \times B \in \tau$, $\forall A \in \tau_1$ et $\forall B \in \tau_2$.

En particulier $\Omega_1 \times \Omega_2 \in \tau$.

Soit $E \in \tau$, $(E^c)_{\omega_1} = (E_{\omega_1})^c \in \tau_2$, donc $E^c \in \tau$.

Soit $(E_i)_i$ une famille dénombrable d'éléments de τ. Il est facile de voir que $(\bigcup_i E_i)_{\omega_1} = \bigcup_i (E_i)_{\omega_1} \in \tau_2$, donc $\bigcup_i E_i \in \tau$.

τ est donc une tribu contenue dans $\tau_1 \otimes \tau_2$ et contenant les rectangles $A \times B$, $A \in \tau_2$, $B \in \tau_2$, donc $\tau = \tau_1 \otimes \tau_2$.

□

Corollaire 6. *Soit f une fonction mesurable de $(\Omega_1 \times \Omega_2, \tau_1 \otimes \tau_2)$ dans $\mathbb{R}$. Pour tout ω_1 de Ω_1, l'application $f_{\omega_1} : \omega_2 \longmapsto f(\omega_1, \omega_2)$ est mesurable de $(\Omega_1 \times \Omega_2)$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. De même l'application $f^{\omega_2} : \omega_1 \longmapsto f(\omega_1, \omega_2)$ pour tout ω_2 de Ω_2 est mesurable de (Ω_1, τ_1) dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.*

©ENS/FAST/UP 2012

Démonstration. Soit $a \in \mathbb{R}$, f étant mesurable, l'ensemble

$$E = \{(\omega_1, \omega_2) \in (\Omega_1 \times \Omega_2), f(\omega_1, \omega_2) > a\} \in \tau_1 \otimes \tau_2$$

D'après la proposition précédente E_{ω_1} et E^{Ω_2} sont mesurables. Or

$$E_{\omega_1} = \{\omega_2 \in \Omega_2, f_{\omega_1}(\omega_2) f(\omega_1, \omega_2) > a\}$$

$$E^{\omega_2} = \{\omega_1 \in \Omega_1, f^{\omega_2}(\omega_1) f(\omega_1, \omega_2) > a\}$$

Il en résulte que f_{ω_1} et f^{ω_2} sont mesurables. □

Théorème 3.1. *(Mesure produit)*
Soit $(\Omega_1, \tau_1, \mu_1)$, $(\Omega_2, \tau_2, \mu_2)$ deux espaces mesurés de mesures μ_1 et μ_2 σ-finies. Alors, il existe une unique mesure μ sur $(\Omega_1 \times \Omega_2, \tau_1 \otimes \tau_2)$ telle que $\forall A_1 \in \tau_1, \forall A_2 \in \tau_2, \mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$. On la note : $\mu = \mu_1 \otimes \mu_2$ et on l'appelle mesure produit de μ_1 et μ_2.

La mesure $\mu = \mu_1 \otimes \mu_2$ vérifie :
$\forall \mathcal{E} \in \tau_1 \otimes \tau_2$,

$$\begin{aligned} \mu_1 \otimes \mu_2(\mathcal{E}) &= \int_{\Omega_1} \mu_2(E_{\omega_1}) d\mu_1(\omega_1) \\ &= \int_{\Omega_2} \mu_1(E^{\omega_2}) d\mu_2(\omega_2) \end{aligned}$$

Corollaire 7. *Soient P_1 et P_2 deux mesures de probabilité sur les espaces probabilisables respectifs, (Ω_1, τ_1), (Ω_2, τ_2). Alors il exite une mesure de probabilité et une seule P sur $(\Omega_1 \times \Omega_2, \tau_1 \otimes \tau_2)$ notée $P = P_1 \otimes P_2$ telle que :*

$$\forall A_1 \in \tau_1, \forall A_2 \in \tau_2, P(A_1 \times A_2) = P_1(A_1) \times P_2(A_2)$$

P_1 et P_2 sont appelées les mesures de probabilité marginales de $P = P_1 \otimes P_2$. En outre P vérifie, $\forall \mathcal{E} \in \tau_1 \otimes \tau_2$,

$$\begin{aligned} P_1 \otimes P_2(\mathcal{E}) &= \int_{\Omega_1} P_2(E_{\omega_1}) dP_1(\omega_1) \\ &= \int_{\Omega_2} P_1(E^{\omega_2}) dP_2(\omega_2) \end{aligned}$$

Théorème 3.2. *(Fubini - Tonelli)*

$(\Omega_1, \tau_1, \mu_1)$ et $(\Omega_2, \tau_2, \mu_2)$ deux espaces mesurés de mesures μ_1 et μ_2 σ-finies. Soit f une fonction mesurable de $(\Omega_1 \times \Omega_2, \tau_1 \otimes \tau_2)$, à valeurs dans $\bar{\mathbb{R}}_+$. Alors les fontions

$$F_1 : \omega_1 \longmapsto \int_{\Omega_2} f(\omega_1, \omega_2) d_{\mu_2}(\omega_2)$$

$$F_2 : \omega_2 \longmapsto \int_{\Omega_1} f(\omega_1, \omega_2) d_{\mu_1}(\omega_1)$$

définies respectivement de ω_1 dans $\bar{\mathbb{R}}_+$ et ω_2 dans $\bar{\mathbb{R}}_+$ sont mesurables et on a :

$$\begin{aligned}\int_{\Omega_1\times\Omega_2} f d(\mu_1\otimes\mu_2) &= \int_{\Omega_1}\{\int_{\Omega_2} f(\omega_1,\omega_2)d_{\mu_2}(\omega_2)\}d\mu_1(\omega_1)\\ &= \int_{\Omega_2}\{\int_{\Omega_1} f(\omega_1,\omega_2)d_{\mu_1}(\omega_1)\}d\mu_2(\omega_2)\end{aligned}$$

Nous énonçons ci-dessous une version plus générale.

Théorème 3.3. *(Fubini)*

(Ω_1,τ_1,μ_1) et (Ω_2,τ_2,μ_2) deux espaces mesurés de mesures μ_1 et μ_2 σ-finies. Soit f une fonction mesurable de $(\Omega_1\times\Omega_2,\tau_1\otimes\tau_2)$, à valeurs dans $(\mathbb{R},\mathcal{B}(\mathbb{R}))$. Pour que f soit $\mu_1\otimes\mu_2$-intégrable, il faut et il suffit que l'une des deux intégrales :

$$I=\int_{\Omega_1}\{\int_{\Omega_2}|f(\omega_1,\omega_2)|d_{\mu_2}(\omega_2)\}d\mu_1(\omega_1)$$

$$J==\int_{\Omega_2}\{\int_{\Omega_1}|f(\omega_1,\omega_2)|d_{\mu_1}(\omega_1)\}d\mu_2(\omega_2)$$

soient finies. En outre , on a :

$$\begin{aligned}\int_{\Omega_1\times\Omega_2} f d(\mu_1\otimes\mu_2) &= \int_{\Omega_1}\{\int_{\Omega_2} f(\omega_1,\omega_2)d_{\mu_2}(\omega_2)\}d\mu_1(\omega_1)\\ &= \int_{\Omega_2}\{\int_{\Omega_1} f(\omega_1,\omega_2)d_{\mu_1}(\omega_1)\}d\mu_2(\omega_2)\end{aligned}$$

Démonstration. Les intégrales I, J et $\int_{\Omega_1\times\Omega_2}|f|d(\mu_1\otimes\mu_2)$ finies ou infinies sont toujours égales en vertu du théorème de Fubini-Tonelli appliqué à $|f|$. Donc, si I ou J est finie, alors, $\int_{\Omega_1\times\Omega_2}|f|d(\mu_1\otimes\mu_2)$ est finie, d'où f est $\mu_1\otimes\mu_2$-intégrable.

□

Cas particulier

On suppose

$$\begin{array}{rlcl} f=f_1\otimes f_2 & :\Omega_1\times\Omega_2 & \longrightarrow & \mathbb{R}\\ & (\omega_1,\omega_2) & \longmapsto & f_1(\omega_1)f_2(\omega_2)\end{array}$$

Corollaire 8. *On suppose que f est de la forme $f=f_1\otimes f_2$. Si f_1 est μ_1-intégrable, f_2, μ_2-intégrable , alors f est $(\mu_1\otimes\mu_2)$-intégrable, et on a :*

$$\int_{\Omega_1\times\Omega_2} f_1\otimes f_2 d(\mu_1\otimes\mu_2)=\{\int_{\Omega_1} f_1 d\mu_1\}\{\int_{\Omega_2} f_2 d\mu_2\}$$

Démonstration.

$$\begin{aligned}\int_{\Omega_1}\{\int_{\Omega_2}|f_1(\omega_1)||f(\omega_2)|d_{\mu_2}(\omega_2)\}d\mu_1(\omega_1) &= \int_{\Omega_1}|f_1(\omega_1)|\{\int_{\Omega_2}|f(\omega_2)|d_{\mu_2}(\omega_2)\}d\mu_1(\omega_1)\\ &= \{\int_{\Omega_1}|f_1(\omega_1)|d\mu_1(\omega_1)\}\{\int_{\Omega_2}|f(\omega_2)|d_{\mu_2}(\omega_2)\}<+\infty\end{aligned}$$

Donc d'après le théorème de Fubini : $\int_{\Omega_1\times\Omega_2} |f_1 \otimes f_2| d(\mu_1 \otimes \mu_2)$ est finie d'où, $f = f_1 \otimes f_2$, est $\mu_1 \otimes \mu_2$-intégrable, et alors,

$$\int_{\Omega_1\times\Omega_2} f_1 \otimes f_2 d(\mu_1 \otimes \mu_2) = \{\int_{\Omega_1} f_1 d\mu_1\}\{\int_{\Omega_2} f_2 d\mu_2\}$$

□

Remarque 3.1. *1. D'après le théorème de Fubini, si $f \in \mathcal{L}^1_{\mu_1\otimes\mu_2}(\Omega_1 \times \Omega_2)$, c'est-à-dire, f intégrable par rapport à la mesure $\mu_1 \otimes \mu_2$, alors $f_{\omega_1} \in \mathcal{L}^1_{\mu_2}(\Omega_2)$ μ_1-presque partout, pour tout $\omega_1 \in \Omega_1$, et la fonction*

$$\omega_1 \longmapsto \int_{\omega_2} f_{\omega_1}(\omega_2) d\mu_2(\omega_2) = \int_{\omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2)$$

est définie μ_1-presque partout sur Ω_1 et μ_1-intégrable sur Ω_1. Un énoncé analogue pour f^{ω_2} en permuttant les rôles de μ_1 et μ_2 est aussi valable.

2. On peut étendre par une démarche analogue à celle suivie jusqu'ici les théorèmes de Fubini au produit de n mesures σ-finies.

©ENS/FAST/UP 2012

Chapitre Quatre

Dépendance et indépendance

4.1 Indépendance d'évènements

4.1.1 Langage des probabilités

Soit (Ω, τ, P) un espace probabilisé .

1. L'ensemble Ω est dit ensemble fondamental. Il représente l'ensemble des résultats possibles d'une expérience aléatoire.
2. Les éléments de τ sont dit évènements. Un évènement est un fait attaché à une expérience aléatoire et susceptible ou non de se produire.
3. Si ω est une épreuve et A un évènement, on dit que l'évènement A se réalise dans l'épreuve ω si et seulement si $\omega \in A$.
4. Le nombre $P(A)$ est la probabilité de l'évènement A.
5. La partie vide ϕ est dite évènement impossible et la partie pleine Ω, est l'évènement certain. Il se réalise dans toute épreuve.
6. Soient A et B deux évènements. On dit que l'évènement A entraiîne l'évènement B si B se réalise dans toute épreuve où A se réalise. Dans ce cas $A \subset B$.
7. Si $A = B$, on dit que les deux évènements sont équivalents.
8. La conjonction de deux évènements A et B est l'événement qui se réalise exactement dans toute épreuve où A et B se réalisent tous deux. C'est l'évènement $A \cap B$.
9. L'évènement $A \cup B$, c'est celui qui se réalise exactement dans toute épreuve où l'un au moins des évènements A, B se réalisent.
10. L'évènement A^c est l'évènement contraire de A.
11. Si $A \cap B = \phi$, on dit que les deux évènements A et B sont incompatibles.
12. Un évènement A tel que $P(A) = 0$, est dit un évènement presque impossible.
13. Un événement A tel que $P(A) = 1$, est dit un évènement presque certain.

4.1.2 Evènements indépendants

Définition 4.1. *Soit (Ω, τ, P) un espace probabilisé. Deux évènements A et B sont dits indépendants si $P(A \cap B) = P(A)P(B)$.*

Remarque 4.1. *1. Si A est un évènement tel que $P(A) = 0$ ou $P(A) = 1$, alors il est indépendent de tout évènement y compris de lui-même.*

©ENS/FAST/UP 2012

2. *Deux évènements incompatibles A et B avec $P(A) > 0$ et $P(B) > 0$ ne sont jamais indépendants.*
3. *En effet, $A \cap B = \phi \Longrightarrow P(A \cap B) = 0$, or $P(A).P(B) \neq 0$.*
4. *Si $P(B) \neq 0$, $P(A/B) = P(A)$, lorsque A et B sont indépendants.*

Proposition 4.1. *Soit A un évènement et $\mathcal{D}_A$ la famille des évènements indépendants de A . Alors $\mathcal{D}_A$ est un système de Dynkin.*

Démonstration. D'après a) de la remarque $\mathcal{D}_A$ contient Ω. Soit B et C éléments de $\mathcal{D}_A$ tels que $B \subset C$.

$$\begin{array}{rcl} B \subset C & \Longrightarrow & A \cap B \subset A \cap C \\ & \Longrightarrow & A \cap C = (A \cap C \ A \cap B) \cup (A \cap B) \end{array}$$

$$P(A \cap C) = P(A \cap C \ A \cap B) + P(A \cap B) \Longrightarrow P(A \cap C) - P(A \cap B) = P[A \cap (C \backslash B)]$$

Or $P(A \cap C) = P(A)P(C)$ et $P(A \cap B) = P(A).P(B)$. Donc,

$$\begin{array}{rcl} P[A \cap (C \backslash B)] & = & P(A).P(C) - P(A).P(B) \\ & = & P(A)[P(C) - P(B)] \\ & = & P(A).P(C \backslash B) \end{array}$$

Ainsi, $P[A \cap (C \backslash B)] = P(A).P(C \backslash B)$, d'où $C \backslash B \in \mathcal{D}_A$. Soit $(A_n)_n$ une famille dénombrable d'évènements deux à deux incompatibles de $\mathcal{D}_A$.

$$\begin{array}{rcl} P[A \cap (\bigcup\limits_n A_n)] & = & P[\bigcup\limits_n (A \cap A_n)] \\ & = & \sum\limits_n P(A \cap A_n) \\ & = & \sum\limits_n P(A).P(A_n) \\ & = & P(A).\sum\limits_n P(A_n) \\ & = & P(A)P(\bigcup\limits_n A_n) \end{array}$$

Ainsi,

$$P[A \cap (\bigcup_n A_n)] = P(A)P(\bigcup_n A_n)$$

Il en résulte que $\mathcal{D}_A$ est un système de Dynkin.

□

Définition 4.2. *Soit $I \subset \mathbb{N}$ et $(A_i)_{i \in I}$ une famille d'évènements. On dit que les évènements $(A_i)_{i \in I}$ sont mutuellement indépendants si,*

$$\forall J fini \subset I, \ P(\bigcap_{j \in J} A_j) = \prod_{j \in J} P(A_j)$$

 ©ENS/FAST/UP 2012

Remarque 4.2. *L'indépendance mutuelle est plus forte que l'indépendance deux à deux. Soient A, B, C les évènements deux à deux indépendants. on a :*

$$P(A \cap B) = P(A).P(B)$$

$$P(B \cap C) = P(B).p(C)$$

$$P(C \cap A) = P(C).P(A)$$

Pour que A, B, et C soient mutuellement indépendants, ils devraient vérifier en plus :

$$P(A \cap B \cap) = P(A).P(B).P(C)$$

Exemple

On lance deux dés et on désigne par A l'évènement " le 1^{er} dé donne un nombre pair ", par l'évènement B l'évènement "le $2^{\text{ème}}$ dé donne un nombre impaire," par C l'évènement "les 2 déc donnent des nombres de même parité."

$\Omega = \{(i,j), 1 \leq i \leq 4, 1 \leq j \leq 6\}$. On a $P(A) = P(B) = P(C) = \frac{1}{2}$ puis $P(A \cap B) = P(B \cap C) = P(A \cap C) = \frac{1}{4}$. Cependant $P(A \cap B \cap C) = 0$ et $P(A).P(B).P(C) \neq 0$

4.2 Indépendance de classes

Définition 4.3. *Soient $\mathcal{C}_1$ et $\mathcal{C}_2$ deux classes d'évènements. On dit que les classes $\mathcal{C}_1$ et $\mathcal{C}_2$ sont indépendantes si : $\forall A_1 \in \mathcal{C}_1$, $\forall A_2 \in \mathcal{C}_2$, A_1 et A_2 sont indépendants, c'est-à-dire $P(A_1 \cap A_2) = P(A_1).P(A_2)$.*

De même soit $I \subset \mathbb{N}$ et $(\mathcal{C}_i)_{i \in I}$ une famille de classes d'évènements. La famille $(\mathcal{C}_i)_{i \in I}$ est dite mutuellement indépendante si :

$\forall J$ fini $\subset I$ et $\forall (\mathcal{A}_j)_{j \in J}$ avec $A_j \in \mathcal{C}_j$.

$$P(\bigcap_{j \in J} A_j) = \prod_{j \in J} P(A_j)$$

En particulier si la famille $(\mathcal{C}_i)_i$ est une famille de tribus, on parle de tribus mutuellement indépendantes.

Proposition 4.2. *Soient $\mathcal{C}_1$, $\mathcal{C}_2$, ..., $\mathcal{C}_n$ des classes indépendantes d'évènements. Si chaque $\mathcal{C}_i$, $1 \leq i \leq n$ est stable par intersection finie, alors les tribus engendrées $\tau(\mathcal{C}_1)$, $\tau(\mathcal{C}_2)$,..., $\tau(\mathcal{C}_n)$ sont indépendantes.*

Démonstration. Nous nous contentons def aire la preuve pour $n = 2$. Soit A_1 fixé dans $\mathcal{C}_1$. Posons

$$\mathcal{F}_{A_1} = \{A_2 \in \tau(\mathcal{C}_2), P(A_1 \cap A_2) = P(A_1).P(A_2)\}$$

$\mathcal{F}_{A_1}$ est la famille des évènements indépendants de A_1, donc d'après la proposition 1, $\mathcal{F}_{A_1}$ est un système de Dynkin. $\mathcal{C}_2 \subset \mathcal{F}_{A_1}$ presque pour tout A_2 élément de $\mathcal{C}_2$, $P(A_1 \cap A_2) = P(A_1).P(A_2)$ à cause de l'indépendance de $\mathcal{C}_1$ et $\mathcal{C}_2$. Mais d'après la proposition ?, Chapître I,

©ENS/FAST/UP 2012

$$\tau(\mathcal{C}_2) = \mathcal{D}(\mathcal{C}_2)$$

où $\mathcal{D}(\mathcal{C}_2)$ est le système de Dynkin engendré par $\mathcal{C}_2$. Il en résulte, $\mathcal{C}_2$ étant inclus dans $\mathcal{F}_{A_1}$ que $\tau(\mathcal{C}_2) \subset \mathcal{F}_{A_1}$ et donc $\tau(\mathcal{C}_2) = \mathcal{F}_{A_1}$. Ainsi on vient de montrer que

$\forall A_1 \in \mathcal{C}_1,\ \forall A_2 \in \tau(\mathcal{C}_2), P(A_1 \cap A_2) = P(A_1).P(A_2)$. Fixons $A_2 \in \tau(\mathcal{C}_2)$ et posons ,

$$\mathcal{F}_{A_2} = \{A_1 \in \tau(\mathcal{C}_\infty), P(A_1 \cap A_2) = P(A_1).P(A_2)\}$$

$\mathcal{F}_{A_2}$ est un système de Dynkin qui contient $\mathcal{C}_1$, et par un raisonnement analogue, on a :

$$\tau(\mathcal{C}_1) = \mathcal{F}_{A_2},\ \forall A_2 \in \tau(\mathcal{C}_2)$$

.

Ainsi, on a établit que :

$$\forall A_1 \in \tau(\mathcal{C}_1), \forall A_2 \in \tau(\mathcal{C}_2), P(A_1 \cap A_2) = P(A_1).P(A_2)$$

□

Corollaire 9. *Si A_1, A_2,..., A_n sont des évènements indépendants, alors toute suite B_1, B_2,..., B_n telle que $B_i = A_i$, où $B_i = A_i^c$ pour chaque i forme aussi des évènements indépendants.*

Démonstration. Les classes $\{A_1\}$, $\{A_2\}$,..., $\{A_n\}$ sont indépendantes et chacun est stable par intersection donc $\tau(\{A_1\})$, $\tau(\{A_2\})$, ..., $\tau(\{A_n\})$ sont des tribus indépendantes. Dans le but de donner un lem connu sous le nom du deuxième lem de Borel Cantelli, rappelons que si $(A_n)_n$ est une suite d'évènements,

$$\limsup_{n\to+\infty} A_n = \bigcap_n \bigcup_{k\geq n} A_k = \{\omega \in \Omega, \omega \text{ réalise une infinité de } A_n\}$$

□

Lemme 4.1. *(Borel Cantelli)*
Soit $(A_n)_n$ une suite d'évènements indépendants telle que $\sum_n P(A_n) = +\infty$.

Alors $P(\limsup_{n\to+\infty} A_n) = 1$, autrement dit ,presque sûrement une infinité d'évènements A_n se réalisent.

Démonstration. Posons $B_{n,m} = \bigcup_{n\leq k\leq m} A_k$, $B_n = \bigcup_{k\geq n} A_k$.

$$P(B_{n,m}) = 1 - P(B_{n,m}^c) = 1 - P(\bigcup_{n\leq k\leq m} A_k^c)$$

Les A_k étant indépendants les A_k^c le sont aussi, d'où,

$$P(B_{n,m}) = 1 - P(\bigcup_{n\leq k\leq m} A_k^c) = 1 - \prod_{k=n} n(1 - P(A_k))$$

En utilisant l'inégalité de convexité,

$$1 - x \leq e^{-x}, 0 < x < 1 \text{ avec } x = P(A_k)$$

on obtient :

$$1 \geq P(B_{n,m}) \geq 1 - \prod_{k=n}^{n} e^{-P(A_k)} = 1 - e^{-\sum_{k=n}^{n} P(A_k)}$$

On fixe n et on fait tendre m vers l'infini, et on obtient grâce au fait que $\sum_n P(A_n) = +\infty$ que :

$$\lim_{n \to +\infty} P(B_{n,m}) = 1$$

D'autre part $\lim_{m \to +\infty} B_{n,m} = B_n$ et puisque $(B_{n,m})_n$ est une suite croissante, on a :

$$P(B_n) = P(\lim_{m \to +\infty} B_{n,m}) = \lim_{m \to +\infty} P(B_{n,m}) = 1, \, \forall n$$

Il en résulte que $P(B_n^c) = 0$ et donc

$$P(\bigcup_n B_n^c) \leq \sum_n P(B_n^c) = 0$$

$P(\bigcup_n B_n^c) = 0$ implique $P(\bigcap_n B_n^c) = 1$, or : $\bigcap_n B_n = \limsup_{n \to +\infty} A_n$, d'où

$$P(\limsup_{n \to +\infty} A_n) = 1$$

.

□

4.3 Indépendance de variables aléatoires

4.3.1 Probabilité conjointes de d variables aléatoires

Définition 4.4. *Soit $(X_1, X_2, ..., X_d)$ un systèmes de d variables aléatoires réelles, on appelle loi de probabilité conjointe du système, la loi de probabilité P_X du vecteur aléatoire $X = (X_1, X_2, ..., X_d)$ à valeurs dans $\mathbb{R}^d$.*

Définition 4.5. *$X = (X_1, X_2, ..., X_d)$ un système de d variables aléatoires réelles, on appelle loi de probabilité marginale P_{X_i} en X_i du système $(X_1, X_2, ..., X_d)$, où du vecteur X, la probabilité image de P_X par la $i^{ème}$ projection $\prod_i$ de $\mathbb{R}$. C'est-à-dire,*

$$\forall B \in \mathcal{B}(\mathbb{R}), P_{X_i}(B) = P_X[\prod_i -1(B)] = P[(\prod_i \circ X)^{-1}(B)]$$

Autrement dit, c'est loi de probabilité de la variable aléatoire X_i.

4.3.2 Variables aléatoires indépendantes

Nous rappelons que si X est une application de (Ω, τ) vers (Ω', τ') où, (Ω, τ) et (Ω', τ') sont deux espaces mesurables, alors :

$$X^{-1}(\tau') = \{x^{-1}(B), B \in \tau'\}$$

est une tribu. C'est la tribu engendrée par X et nous la notons $\tau(X)$. Si en plus X est mesurable, alors $\tau(X)$ est une sous-tribu de τ.

©ENS/FAST/UP 2012

Définition 4.6. *Soit $I\mathbb{N}$ et $(X_i)_{i\in I}$ une famille de variables aléatoires définies sur un espace probabilisé (Ω,τ,P). On suppose que chaque X_i est à valeurs dans un espace mesurable (Ω_i,τ_i). On dit que la famille $(X_i)_{i\in J}$ est une famille de variables aléatoires indépendantes si la famille de tribus $(\tau(X_i))_{i\in J}$ est mutuellement indépendante, autrement dit,*

$$\forall J fini \subset I,\ \forall B_j \in \tau_j,\ \forall j \in J, P(\bigcap_{j\in J} X^j{-}1(B_j) \forall j \in J) = \prod_{j\in J} P(X_j{-}1(B_j))$$

Remarque 4.3. *Cette définition s'applique aussi à une famille où $(X_i)_{i\in I}$, où X_i pouvait être des variables aléatoires réelles, d'autres des vecteurs aléatoires.*

Proposition 4.3. *Soit un vecteur aléatoire $X=(X_1,X_2,...,X_d)$ à valeurs dans $\mathbb{R}^d$. Les composantes $X_1,X_2,...,X_d$ sont indépendantes si et seulement la loi de probabilité P_X de X est le produit de ses lois de probabilité marginales. C'est-à-dire,*

$$P_X = P_{(X_1,X_2,...,X_d)} = P_{X_1}\otimes P_{X_2}\otimes ...\otimes P_{X_d}$$

Démonstration. Soit $B\in\prod_{i=1}^d \mathcal{B}(\mathbb{R})$ avec $\mathcal{B}_i(\mathbb{R})=\mathcal{B}(\mathbb{R})$, $\forall i$.

$$B = B_1\times B_2\times ...\times B_d \text{ avec } B_i\in\mathcal{B}(\mathbb{R})\, i=1,2,...,d$$

et

$$B\in \underbrace{\mathcal{B}(\mathbb{R})\otimes\mathcal{B}(\mathbb{R})\otimes ...\otimes\mathcal{B}(\mathbb{R})}\, dfois = \mathcal{B}(\mathbb{R}^d)$$

$$\begin{aligned} P_{(X_1,X_2,...,X_d)}(B) &= P((X_1,X_2,...,X_d)\in B_1\times B_2\times ...\times B_d) \\ &= P(X_i\in B_i,\ \forall i\in\{1,2,...,d\}) \\ &= P(\bigcap_{i=1}^d X_i^{-1}(B_i)) \end{aligned}$$

Si les $X_1,X_2,...,X_d$ sont indépendantes, alors,

$$\begin{aligned} P(\bigcap_{i=1} dX_i^{-1}(B_i)) &= \prod_{i=1} dP(X_i^{-1}(B_i)) \\ &= \prod_{i=1} dP_{X_i^{-1}}(B_i) \\ &= (P_{X_1}\otimes P_{X_2}\otimes ...\otimes P_{X_d})(B) \end{aligned}$$

Ainsi :

$$P_{(X_1,X_2,...,X_d)}(B) = (P_{X_1}\otimes P_{X_2}\otimes ...\otimes P_{X_d})(B)$$

Donc $P_{(X_1,X_2,...,X_d)}$ coincide avec $P_{X_1}\otimes P_{X_2}\otimes ...\otimes P_{X_d}$ sur $\prod_{i=1}^d\mathcal{B}(\mathbb{R})$ qui engendre $\mathcal{B}(\mathbb{R}^d)$. Par unicité de la mesure produit, $P_{(X_1,X_2,...,X_d)}$ coincidant avec $P_{X_1}\otimes P_{X_2}\otimes ...\otimes P_{X_d}$ sur $\mathcal{B}(\mathbb{R}^d)$. Réciproquement supposons que $P_{(X_1,X_2,...,X_d)} = P_{X_1}\otimes P_{X_2}\otimes ...\otimes P_{X_d}$ sur $\mathcal{B}(\mathbb{R}^d)$.

□

4.3.3 Fonction de répartition d'un vecteur aléatoire

Définition 4.7. *Soit (Ω, τ, P) un espace probabilisé et $X = (X_1, X_2, ..., X_d)$ un vecteur aléatoire sur (Ω, τ) à valeurs dans $\mathcal{B}(\mathbb{R}^d)$. On applle fonction répartition de X la fonction F définie par :*

$$F(x_1, x_2, ..., x_d) = P(X_1 \leq x_1, X_2 \leq x_2, ..., X_d \leq x_d),\ \forall (x_1, x_2, ..., x_d) \in \mathbb{R}^d$$

Remarque 4.4. *1. La fonction de répartition F peut encore se mettre sous la forme :*

$$\begin{aligned} F(x_1, x_2, ..., x_d) &= P(X_1 \leq x_1, X_2 \leq x_2, ..., X_d \leq x_d) \\ &= P(]-\infty; x_1] \times]-\infty; x_2] \times ... \times]-\infty; x_d]) \end{aligned}$$

pour tout $(x_1, x_2, ..., x_d) \in \mathbb{R}^d$.

2. F est croissante et continue à droite en chacune de ses variables.

Proposition 4.4. *Soit $X = (x_1, x_2, ..., x_d)$ un vecteur aléatoire à valeurs dans $\mathbb{R}^d$ de fonction de répartition F. Soient $F_1, F_2, ..., F_d$ les fonctions de répatitions marginales des variables aléatoires $X_1, X_2, ..., X_n$. Alors, on a :*

1.

$$F_1(x_1) = P(X_1 \leq x_1) = \lim_{(x_2,...,x_d) \to (+\infty,...,+\infty)} F(x_1, x_2, ..., x_d)$$

$$F_2(x_2) = P(X_2 \leq x_2) = \lim_{(x_1,x_3,...,x_d) \to (+\infty,...,+\infty)} F(x_1, x_2, ..., x_d)$$

$$F_d(x_d) = P(X_d \leq x_d) = \lim_{(x_1,x_2,...,x_{d-1}) \to (+\infty,...,+\infty)} F(x_1, x_2, ..., x_d)$$

2. Chaque F_i est une fonction croissante et continue à droite.

Démonstration. 1. Nous limiterons la preuve au cas de deux variables aléatoires X_1, X_2 et à la fonction F_1. La généralisation à un nombre quelconque de variables aléatoires est une simple adaptation du cas de deux variables.

$$\begin{aligned} F_1(x_1) &= P_{X_1}(]-\infty, x_1]) \\ &= P(X_1^{-1}(]-\infty, x_1])) \\ &= P[(\prod_1 \circ X)^{-1}(]-\infty, x_1])] \\ &= P(X_1^{-1}(\prod_1^{-1}(]-\infty, x_1]))) \\ &= P(\prod_1^{-1}(]-\infty, x_1])) \end{aligned}$$

$\prod_1$ étant la première projection de $\mathbb{R}^2$. $\prod_1^{-1}(]-\infty, x_1]) =]-\infty, x_1] \times \mathbb{R}$, ce qui implique que :

$$F_1(x_1) = P_X(]-\infty, x_1] \times \mathbb{R})$$

©ENS/FAST/UP 2012

. Par ailleurs,

$$F(x_1, x_2) = P_X(]-\infty, x_1]\times]-\infty, x_2])$$

F étant croissante par rapport à chaque variable l'application $x_2 \longmapsto F(x_1, x_2)$ est croissante $\forall x_1 in \mathbb{R}$, et donc $\lim_{x_2\to+\infty} F(x_1, x_2)$, existe et par conséquent :

$$\lim_{x_2\to+\infty} F(x_1, x_2) = \lim_{n\to+\infty} F(x_1, n) = \lim_{n\to+\infty} P_X(]-\infty, x_1]\times]-\infty, n])$$

. Or la suite $(B_n)_n$ telle que $B_n =]-\infty, x_1]\times]-\infty, n]$ est une suite croissante et $\lim_{n\to+\infty} B_n =]-\infty, x_1]\times\mathbb{R}$. En utilisant la continuité monotone séquentielle de P_X, on a

$$\lim_{n\to+\infty} P_X(]-\infty, x_1]\times]-\infty, n]) = P_X(]-\infty, x_1]\times\mathbb{R})$$

Ainsi $\lim_{x_2\to+\infty} F(x_1, x_2) = P_X(]-\infty, x_1]\times\mathbb{R}) = F_1(x_1)$.

2. La fonction de répartition F étant croissante et continue à droite en chacun de ses variables, chaque F_i est une fonction croissante et continue à droite.

□

Proposition 4.5. *Soit $X = (X_1, X_2, ..., X_d)$ un vecteur aléatoire à valeurs dans $\mathbb{R}^d$. Les variables marginales $X_1, X_2, ..., X_d$ sont indépendants si et seulement si la fonction répartition F de X vérifie*

$$F(x_1, x_2, ..., x_d) = F_1(x_1).F_2(x_2)...F_d(x_d)\ \forall (x_1, x_2, ..., x_d) \in \mathbb{R}^d$$

où $F_1(x_1)$, $F_2(x_2)$, ..., $F_d(x_d)$ sont des fonctions de répartitions marginales des variables $X_1, X_2, ..., X_d$.

Démonstration. $X_1, X_2, ..., X_d$ sont indépendants si et seulement si :

$$P_(X_1, X_2, ..., X_d) = P_{x_1} \otimes P_{x_2} \otimes ... \otimes P_{x_d} \text{ sur } \mathcal{B}(\mathbb{R}^d)$$

.

Supposons $X_1, X_2, ..., X_d$ indépendantes. Alors ,
$\forall (x_1, x_2, ..., x_d) \in \mathbb{R}^d$, $B =]-\infty, x_1]\times]-\infty, x_2]\times ...\times]-\infty, x_d]$ appartient à $\mathcal{B}(\mathbb{R}^d)$ et

$$\begin{aligned} P_(X_1, X_2, ..., X_d)(B) &= P_X(]-\infty, x_1]\times]-\infty, x_2]\times ...\times]-\infty, x_d]) \\ &= F(x_1, x_2, ..., x_d) \end{aligned}$$

$$(P_{x_1} \otimes P_{x_2} \otimes ... \otimes P_{x_d})(B) = P_{X_1}(]-\infty, x_1]).P_{X_2}(]-\infty, x_2])...P_{X_d}(]-\infty, x_d])$$

par définition de mesure produit puisque :

$$B =]-\infty, x_1]\times]-\infty, x_2]\times ...\times]-\infty, x_d] \in \prod_{i=1}^{d} \mathcal{B}(\mathbb{R}^d)$$

Il résulte de (a) et (b) que :

$$\begin{aligned} F(x_1, x_2, ..., x_d) &= P_{X_1}(]-\infty, x_1]).P_{X_2}(]-\infty, x_2])...P_{X_d}(]-\infty, x_d]) \\ &= F_1(x_1).F_2(x_2)....F_d(x_d) \end{aligned}$$

$\forall (x_1, x_2, ..., x_d) \in \mathbb{R}^d$.

Réciproquement supposons que,

$$\begin{aligned} F(x_1, x_2, ..., x_d) &= F_1(x_1).F_2(x_2)....F_d(x_d) \\ &= P_{X_1}(]-\infty, x_1]).P_{X_2}(]-\infty, x_2])...P_{X_d}(]-\infty, x_d]) \end{aligned}$$

$\forall (x_1, x_2, ..., x_d) \in \mathbb{R}^d$.

$$B =]-\infty, x_1] \times]-\infty, x_2] \times ... \times]-\infty, x_d] \in \prod_{i=1}^{d} \mathcal{B}(\mathbb{R}^d)$$

alors

$$P_{X_1}(]-\infty, x_1]).P_{X_2}(]-\infty, x_2])...P_{X_d}(]-\infty, x_d]) = (P_{x_1} \otimes P_{x_2} \otimes ... \otimes P_{x_d})(B)$$

Par ailleurs, $F(x_1, x_2, ..., x_d) = P_{(X_1, X_2, ..., X_d)}(B)$, d'où

$P_{(X_1, X_2, ..., X_d)} = P_{x_1} \otimes P_{x_2} \otimes ... \otimes P_{x_d}$. Ce qui achève la preuve de la proposition.

□

Proposition 4.6. *Soit (Ω, τ, P) un espace prabilisé et $X = (X_1, X_2, ..., X_d)$ un vecteur aléatoire discret de (Ω, τ) dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Alors les composantes $X_1, X_2, ..., X_d$ sont indépendantes si et seulement si :*

$$\forall (x_1, x_2, ..., x_d) \in X(\Omega), P(X_1 = x_1, X_2 = x_2, ..., X_d = x_d) = P(X_1 = x_1).P(X_2 = x_2)...P(X_d = x_d)$$

Démonstration. Supposons $X_1, X_2, ..., X_d$ indépendantes,

$$P_{(X_1, X_2, ..., X_d)}(B) = (P_{x_1} \otimes P_{x_2} \otimes ... \otimes P_{x_d})(B), \forall B \in \mathcal{B}(\mathbb{R}^d)$$

En particulier si $B = \prod_{i=1}^{d} \{x_i\} \in \prod_{i=1}^{d} \mathcal{B}(\mathbb{R}^d)$. On a :

$$\begin{aligned} (P_{X_1} \otimes P_{X_2} \otimes ... \otimes P_{X_d})(\prod_{i=1}^{d} \{x_i\}) &= \prod_{i=1}^{d} P_{X_i}(\{x_i\}) \\ &= P(X_1 = x_1).P(X_2 = x_2)...P(X_d = x_d) \end{aligned}$$

et

$$P_{(X_1, X_2, ..., X_d)}(\prod_{i=1}^{d} \{x_i\}) = P(X_1 = x_1, X_2 = x_2, ..., X_d = x_d)$$

d'où l'égalité.

©ENS/FAST/UP 2012

Réciproquement , supposons l'égalité et montrons que :

$$P_{(X_1,X_2,...,X_d)} = P_{x_1} \otimes P_{x_2} \otimes ... \otimes P_{x_d}$$

Soit $B \in \prod_{i=1}^{d} \mathcal{B}(\mathbb{R}^d)$,

$$P_{(X_1,X_2,...,X_d)}(B) = P(X_1 \in B_1 \cap X_1(\Omega), X_2 \in B_2 \cap X_2(\Omega), ..., X_d \in B_d \cap X_d(\Omega))$$

avec $B = \prod_{i=1}^{d} B_i$

Ainsi,

$$\begin{aligned}
P_{(X_1,X_2,...,X_d)}(B) &= \sum_{x_1 \in B_1 \cap X_1(\Omega)} \sum_{x_2 \in B_2 \cap X_2(\Omega)} ... \sum_{x_d \in B_d \cap X_d(\Omega)} P(X_1 = x_1, X_2 = x_2, ..., X_d = x_d) \\
&= \sum_{(x_1,x_2,...,x_d) \in B \cap X(\Omega)} P(X_1 = x_1, X_2 = x_2, ..., X_d = x_d) \\
&= \sum_{(x_1,x_2,...,x_d) \in B \cap X(\Omega)} P(X_1 = x_1).P(X_2 = x_2)...P(X_d = x_d) \\
&= (\sum_{x_1 \in B_1 \cap X_1(\Omega)} P(X_1 = x_1))(\sum_{x_2 \in B_2 \cap X_2(\Omega)} P(X_2 = x_2))...(\sum_{x_d \in B_d \cap X_d(\Omega)} P(X_d = x_d)) \\
&= P(X_1 \in B_1).P(X_2 \in B_2)...P(X_d \in B_d) \\
&= (P_{X_1} \otimes P_{X_2} \otimes ... \otimes P_{X_d})(B)
\end{aligned}$$

$P_{(X_1,X_2,...,X_d)}$ et $P_{X_1} \otimes P_{X_2} \otimes ... \otimes P_{X_d}$ coincident sur $\prod_{i=1} d\mathcal{B}(\mathbb{R}^d)$, d'où par unicité de la mesure produit, on a :

$$P_{(X_1,X_2,...,X_d)} = P_{X_1} \otimes P_{X_2} \otimes ... \otimes P_{X_d} \text{ sur } \mathcal{B}(\mathbb{R}^d)$$

Ce qui prouve l'indépendance de $X_1, X_2, ..., X_d$.

□

Proposition 4.7. *Soit (Ω, τ, P) un espace probabilisé et $X_1, X_2, ..., X_n$ n variables aléatoires réelles sur (Ω, τ, P). Si les variables aléatoires $X_1, X_2, ..., X_n$ sont indépendantes et intégrables, c'est-à-dire*

$E(X_i) < \infty \ \forall i \in \{1, 2, ..., n\}$, alors $E(X_1, X_2, ..., X_n) < \infty$, et on a :

$$E(X_1, X_2, ..., X_n) = E(X_1)E(X_2)...E(X_n)$$

Démonstration. En utilisant le théorème de transfert, on a :

$$\begin{aligned} E(|X_1.X_2...X_n|) &= \int_\Omega |X_1.X_2...X_n| dP \\ &= \int_{\mathbb{R}^n} |x_1||x_2|...|x_n| dP_{(X_1,X_2,...,X_n)} \\ &= \int_{\mathbb{R}^n} |x_1||x_2|...|x_n| d(P_{X_1} \otimes P_{X_2} \otimes ... \otimes P_{X_d}) \\ &= \prod_{i=1}^{n} \int_{\mathbb{R}} |x_i| dP_{X_i} \\ &= \prod_{i=1}^{n} E(|x_i|) < \infty \end{aligned}$$

Car $E(X_i) < \infty \ \forall i \in \{1, 2, .., n\}$.
Ainsi, $E(|X_1.X_2...X_n|) < \infty$ donc $E(X_1.X_2...X_n) < \infty$. En outre,

$$\begin{aligned} E(X_1.X_2...X_n) &= \int_{\mathbb{R}^n} x_1.x_2...x_n d(P_{X_1} \otimes P_{X_2} \otimes ... \otimes P_{X_d}) \\ &= \prod_{i=1}^{n} \int_{\mathbb{R}} x_i dP_{X_i} \\ &= \prod_{i=1}^{n} E(X_i) \end{aligned}$$

□

Remarque 4.5. *La réciproque de la proposition est fausse. $E(X_1X_2)$ peut être égale à $E(X_1).E(X_2)$ sans que X_1 et X_2 ne soient indépendantes.*

4.4 Dépendance de variables aléatoires

Nous venions de voir que si X et Y sont indépendantes et intégrables, $E(XY) = E(X).E(Y)$ et que la réciproque est fausse. Ceci nous emmène à nous intéresser à l'étude des propriétés de la quantité $E(XY) - E(X).E(Y)$.

4.4.1 Covariance

Définition 4.8. *Soient X et Y deux variables aléatoires telles que $E(X^2) < \infty$ et $E(Y^2) < \infty$. Alors la quantité :*

$$cov(X, Y) = E[(X - E(X)(Y - E(Y))]$$

est bien définie et est appelée covariance de X et Y.

Démonstration. Pour légitimer la définition, il nous faut montrer que :

$$E[(X - E(X))(Y - E(Y))] < \infty$$

©ENS/FAST/UP 2012

$$E(|X|) = \int_{\{|x|\leq 1\}} |X| dP + \int_{\{|x|>1\}} |X| dP \leq 1 + \int_{\{|x|>1\}} X^2 dP$$

Donc, $E(|X|) \leq 1 + E(X^2) < \infty$. De même, $E(|Y|) \leq 1 + E(Y^2) < \infty$. Il en résulte que $E(X) < \infty$ et $E(Y) < \infty$. Par ailleurs,

$$(X - E(X))(Y - E(Y)) \leq (X - E(X))^2 + (Y - E(Y))^2$$

, donc

$$E[(X - E(X))(Y - E(Y))] \leq E[(X - E(X))^2] + E[(Y - E(Y))^2]$$

Or

$$E[(X - E(X))^2] = E[X^2 - 2XE(X) + (E(X))^2] = E(x^2) - (E(X))^2$$

De même pour Y on a :

$$E[(Y - E(Y))^2] \leq E(Y^2) - (E(Y))^2$$

En prenant en compte (a) et (b) on a :

$$E[(X - E(X))(Y - E(Y))] < \infty$$

□

Proposition 4.8. *(Formule de Koenig)*
Soient X et Y deux variables aléatoires telles que $E(X^2) < \infty$ et $E(Y^2) < \infty$. Alors $cov(X, Y) = E(XY) - E(X)E(Y)$.

Démonstration.

$$(X - E(X))(Y - E(Y)) = XY - XE(Y) - YE(X) + E(X)E(Y)$$

$$cov(X, Y) = E(XY) - E(X)E(Y) - E(Y)E(X) + E(X)E(Y)$$

d'où $cov(X, Y) = E(XY) - E(X)E(Y)$.

□

Définition 4.9. *On dit que deux variables aléatoires X et Y telles que $E(X^2) < \infty$ et $E(Y^2) < \infty$ sont non-corrélées si $cov(X, Y) = 0$. Ce équivaut à $E(XY) = E(X)E(Y)$.*

Remarque 4.6. *Deux variables aléatoires X et Y indépendantes et intégrables sont non-corrélées même si elles ne sont pas de carrés intégrables.*

Définition 4.10. *(Variance)*
Soit X une variable aléatoire de carré intégrable. On appelle variance de X la quantité :

$$VarX = E[(X - E(X))^2] = cov(X, X)$$

$$\sigma(X) = \sqrt{VarX} \text{ est l'écart-type.}$$

Proposition 4.9. *Soit $X_1, X_2, ..., X_n$ n variables aléatoires telles que $E(X_i^2) < \infty \forall i \in \{1, 2, .., n\}$. Alors on a :*

$$\begin{aligned} Var(\sum_{k=1}^{n} X_k) &= \sum_{i,j=1}^{n} cov(X_i, X_j) \\ &= \sum_{i=1}^{n} VarX_i + \sum_{\substack{0<i<n \\ i\neq j}}^{n} cov(X_i, X_j) \end{aligned}$$

En particulier si les variables sont deux à deux non-corrélées, alors on a :

$$Var(\sum_{k=1}^{n} X_k) = \sum_{k=1}^{n} X_k$$

Démonstration. Il suffit de constater que l'application

$$(X, Y) \longmapsto cov(X, Y)$$

est une forme bilinéaire sur l'espace vectoriel des variables aléatoires de carré intégrable. □

Définition 4.11. *(Coefficient de corrélation)*
*Soient X et Y deux variables aléatoires telles que $E(X^2) < \infty$ et $E(Y^2) < \infty$. Supposons que $\sigma(X)\sigma(Y) > 0$. On appelle coefficient de corrélation du couple * (X, Y), le nombre :*

$$r(X, Y) = \frac{cov(X, Y)}{\sigma(X).\sigma(Y)} = E[\frac{(X - E(X))}{\sigma(X)}.\frac{Y - E(Y))}{\sigma Y}]$$

Remarque 4.7. *On peut vérifier facilement que si $ac \neq 0$, on a :*

$$r(aX + b, cY + d) = sgn(ac)r(X, Y)$$

Ainsi, si $ac > 0$, r est à la fois invariant à la fois par changement d'origine et par changement d'échelle sur les axes Ox et Oy. En conséquence, on ne modifie pas la valeur de $r(X, Y)$ en supposant dans les calculs, les valeurs marginales X et Y centrées réduites et dans ce cas $r(XY) = E(XY)$.

Propriété 4.1.

$$|r(X, Y)| \leq 1$$

Démonstration. On suppose X et Y centrées réduites ; pour tout □

Saut de PAGE 54

En multipliant la $2^{\text{ème}}$ équation par $-E(X)$, on a :

$$\begin{cases} -2E(XY) + 2aE(X^2) + 2bE(X) &= 0\,(*) \\ 2E(Y)E(X) - 2aE^2(X) - 2bE(X) &= 0\,(**) \end{cases}$$

En additionnant $(*)$ et $(**)$, on a :

$$2E(X)E(Y) - 2E(XY) + 2a(E(X^2) - E^2(X)) = 0$$

 ©ENS/FAST/UP 2012

On en tire que :

$$a = \frac{cov(X,Y)}{Var(X)} = r(X,Y).\frac{\sigma(Y)}{\sigma(X)} = \alpha$$

$$b = E(Y) - aE(X) = \beta$$

4.5 Convolution

Dans cette section (E,τ) est un espace mesurable avec $E = \mathbb{R}^d$ ou $\mathbb{Z}^d$ $(d \geq 1)$ et $\tau = \mathcal{B}(\mathbb{R}^d)$ ou $\mathcal{P}(\mathbb{Z}^d)$ respectivement. Dans les deux cas la tribu τ est invariante par toute translation t_x de vecteur x élément de E.

Définition 4.12. *Soient μ_1 et μ_2 deux mesures finies sur (E,τ). Le produit de convolution de μ_1 et μ_2 est la mesure :*

$$\mu_1 * \mu_2 = (\mu_1 \otimes \mu_2) \circ S^{-1}$$

où $S : E^2 \longmapsto E$, $(x,y) \longmapsto x+y$ est l'addition sur

$\mu_1 * \mu_2$ est une mesure finie sur (E,τ) . En particulier, si μ_1 et μ_2 sont des probabilités sur (E,τ) , $\mu_1 * \mu_2$ est aussi une prabalité sur (E,τ).

Proposition 4.10. *Pour tout $B \in \tau$.*

$$\mu_1 * \mu_2(B) = \int_E \mu_2(B-x)d\mu_1(x) = \int_E \mu_1(B-y)d\mu_2$$

où $B - x = \{z - x, z \in B\} = t_{-x}(B)$.

Démonstration. Posons $A = S^{-1}(B) = \{(x,y) \in E^2, x+y \in B\}$, $A_x = \{y \in E, (x,y) \in A\}$.

$$\begin{aligned} A_x &= \{y \in E, x+y \in B\} \\ &= \{y \in E, \exists z \in B, y = z - x\} \\ &= \{z - x, z \in B\} \\ &= B - x \end{aligned}$$

Ainsi :

$$\begin{aligned} \mu_1 * \mu_2(B) &= \mu_1 \otimes \mu_2(S^{-1}(B)) \\ &= \mu_1 \otimes \mu_2(A) \\ &= \int_E \mu_2(A_x)d\mu_1(x) \\ &= \int_E \mu_2(B-x)d\mu_1(x) \end{aligned}$$

□

La deuxième égalité s'obtient en échangeant les rôles de x et y.

Proposition 4.11. *Soient X et Y deux variables aléatoires indépendantes à valeurs dans (E, τ). Alors la loi de $X + Y$ est le produit de convolution des lois X et Y.*

$$X et Y \text{ indépendantes } \Longrightarrow P_{X+Y} = P_X * P_Y$$

Démonstration. Puisque $X + Y = S \circ (X, Y)$, $P_{X+Y} = P_{S\circ(X,Y)} = P_{(X,Y)} \circ S^{-1}$, X et Y étant indépendantes $P_{(X,Y)} = P_X \otimes P_Y$. D'où

$$P_{X+Y} = P_X \otimes P_Y \circ S^{-1} = P_X * P_Y$$

.

□

Proposition 4.12. *Soit h une fonction de E dans $\bar{\mathbb{R}}_+$, positive et mesurable par rapport à (E, τ). Alors, on a :*

1. $\int_E h d(\mu_1 * \mu_2) = \int_{E^2} h(x + y) d(\mu_1 \otimes \mu_2)(x, y)$.
   En outre, si h est à valeurs dans $\mathbb{R}$.
2. $h \in \mathcal{L}^1_{\mathbb{R}}(\mu_1 \otimes \mu_2) \Longleftrightarrow h \circ S \in \mathcal{L}^1_{\mathbb{R}}(\mu_1 \otimes \mu_2)$

Démonstration. 1. $\mu_1 * \mu_2$ est la mesure image de $\mu_1 \otimes \mu_2$ puis , on applique donc le théorème de transfert à $S : E \times E \longrightarrow E$.

$$\begin{aligned} \int_{E^2} h(x + y) d(\mu_1 \otimes \mu_2)(x, y) &= \int_{E^2} h \circ S d(\mu_1 \otimes \mu_2)(x, y) && (4.1) \\ &= \int_{E^2} h d(\mu_1 \otimes \mu_2 \circ S^{-1})(x, y) && (4.2) \\ &= \int_{E^2} h d(\mu_1 \otimes \mu_2) && (4.3) \end{aligned}$$

2. $\int_{E^2} h d(\mu_1 \otimes \mu_2) = \int_{E^2} |h|(x + y) d(\mu_1 \otimes \mu_2)(x, y)$

$h \in \mathcal{L}^1_{\mathbb{R}}(\mu_1 \otimes \mu_2)$ si et seulement si la première intégrale est finie, ce qui équivaut à la finitude de la deuxième intégrale qui équivaut à son tour à $h \circ S \in \mathcal{L}^1_{\mathbb{R}}(\mu_1 \otimes \mu_2)$. □

Proposition 4.13. *Si μ_1 et μ_2 sont des mesures finies de densités respectives f et g par rapport à la mesure de Lebesgue λ sur $\mathbb{R}^d$, alors $\mu_1 * \mu_2$ est la mesure de densité h définie $\lambda - pp$ par :*

$$h(y) = \int_{\mathbb{R}^d} f(x) g(y - x) d\lambda(x) = \int_{\mathbb{R}^d} f(y - x) g(x) d\lambda(x)$$

Démonstration. Soit $B \in \mathcal{B}(\mathbb{R}^d)$

$$\begin{aligned} \mu_1 * \mu_2(B) &= \int_{\mathbb{R}^d} \mu_2(B - x) f(x) d\lambda(x) \\ &= \int_{\mathbb{R}^d} \{ \int_{\mathbb{R}^d} 1_{B-x}(u) g(u) d\lambda(u) \} f(x) d\lambda(x) \end{aligned}$$

Faisons le changement de variable $y = x + u$ à x fixé, dans l'intégrale intérieure, on a :

$$\int_{\mathbb{R}^d} 1_{B-x}(u) g(u) d\lambda(u) = \int_{\mathbb{R}^d} 1_{B-x}(y - x) g(y - x) d\lambda(y)$$

Or $1_{B-x}(u) = 1_B(x+u)$, donc,

$$\int_{\mathbb{R}^d} 1_{B-x}(y-x)g(y-x)d\lambda(y) = \int_{\mathbb{R}^d} 1_B(y)g(y-x)d\lambda(y)$$

Ainsi en exploitant l'égalité ci-dessus et le théorème de Fubini-Tonelli, on a :

$$\begin{aligned}\mu_1 * \mu_2(B) &= \int_{\mathbb{R}^d} 1_B(y)\{\int_{\mathbb{R}^d} f(x)g(y-x)d\lambda(x)\}f(x)d\lambda(y) \\ &= \int_B \{\int_{\mathbb{R}^d} f(x)g(y-x)d\lambda(x)\}d\lambda(y)\end{aligned}$$

$\forall B \in \mathcal{B}(\mathbb{R}^d)$. Il en résulte que : $\mu_1 * \mu_2(B) = \int_B h(y)d\lambda(y)\ \forall B \in \mathcal{B}(\mathbb{R}^d)$, avec :

$$h(y) = \int_{\mathbb{R}^d} f(x)g(y-x)d\lambda(x)$$

D'où h est la densité définie $\lambda - pp$ de $\mu_1 * \mu_2$. On montre de façon analogue la deuxième égalité en permutant les rôles de μ_1 et μ_2 dans le calcul.

□

©ENS/FAST/UP 2012

Lois de probabilité absolument continues - Lois de probabilité conditionnelle

5.1 Lois de probabilité absolument continues

Définition 5.1. *Soit (X,Y) un couple de variables aléatoires et $P_(X,Y)$ sa loi conjointe. On dit que la $P_(X,Y)$ est absolument continue par rapport à la mesure de Lebesgue λ sur $(\mathbb{R}^2,\mathcal{B}(\mathbb{R}^2))$ s'il existe une fonction mesurable positive $f:(\mathbb{R}^2,\mathcal{B}(\mathbb{R}^2))\longrightarrow(\mathbb{R}^+,\mathcal{B}(\mathbb{R}^2))$ telle que :*

$$P_{(X,Y)}(B)=\int_B f(x,y)dxdy=\int_{\mathbb{R}^2} f(x,y)1_B(x,y)dxdy$$

La fonction f est appelée densité de probabilité conjointe de la loi de (X,Y). On la note aussi $f_{X,Y}(x,y)$. On dit aussi que le couple (X,Y) est absolument continue de densité conjointe f.

Proposition 5.1. *Soit f une densité de probabilité conjointe, f vérifie :*

1. $f\geq 0$
2. $\int_{\mathbb{R}^2} f(x,y)dxdy=1$
3. *La fonction de répartition conjointe $F=F_{X,Y}$ est représentée par :*

$$F(x,y)=\int_{]-\infty,x]\times]-\infty,y]} f(u,v)dudv$$

4. *Si f est continue au point (x_0,y_0), alors,*

$$f(x_0,y_0)=\frac{\partial^2 F(x,y)}{\partial x\partial y}\ \text{ au point } (x,y)=(x_0,y_0)$$

Proposition 5.2. *Soit (X,Y) un couple de variables aléatoires absolument continu de densité conjointe f, alors X et Y sont absolument continues et les fonctions densité marginales f_X, f_Y sont données par*

$$f_X(x)=\int_{\mathbb{R}} f(x,y)dy\,pp$$

$$f_Y(y)=\int_{\mathbb{R}} f(x,y)dx\,pp$$

©ENS/FAST/UP 2012

Démonstration.

$$\begin{aligned} F_1(x) &= \lim_{y\to+\infty} F(x,y) \\ &= F(x,+\infty) \\ &= \int_{]-\infty,x]\times\mathbb{R}} f(u,v)dudv \end{aligned}$$

En utilisant Fubini-Tonelli, on a :

$$F_1(x) = \int_{]-\infty,x]} \{\int_{\mathbb{R}} f(u,v)dv\}du \ (1)$$

F_1 étant absolument continue, on a :

$$F_1(x) = \int_{]-\infty,x]} f_X(u)du \ (2)$$

$$(1)\text{et}(2) \Longrightarrow f_X(u) = \int_{\mathbb{R}} f(u,v)dv \ pp.$$

En inversant les rôles de x et de y, on montre de façon analogue que :

$$f_Y(v) = \int_{\mathbb{R}} f(u,v)du \ pp.$$

□

Proposition 5.3. *Soit (X,Y) un couple de variables aléatoires, absolument continue et de densité conjointe f. Soient f_X, f_Y les densités marginales de X et de Y respectivement. Alors, les deux propriétés suivantes sont équivalentes.*

1. *X et Y sont indépendantes.*
2. *Pour tout $(x,y) \in \mathbb{R}^2$, on a :*

$$f(x,y) = f_X(x).f_Y(y) \ pp$$

Démonstration. Sens 1) $\Longrightarrow$ 2).

Supposons X et Y indépendantes, alors en notant F la fonction de répartition conjointe de (X,Y), F_X, F_Y les fonctions de répartitions marginales, on a :

$$F(x,y) = F_X(x).F_Y(y)$$

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} = \frac{\partial F_X(x)}{\partial x}.\frac{\partial F_Y(y)}{\partial y} = f_X(x).f_Y(y) \ pp$$

Sens 2) $\Longrightarrow$ 1).
Supposons $f(x,y) = f_X(x).f_Y(y)$ *pp* pour tout $(x,y) \in \mathbb{R}^2$. Or alors,

$$\begin{aligned} F(x,y) &= \int_{]-\infty,x]\times]-\infty,y]} f_X(u)f_Y(v)dudv \\ &= (\int_{]-\infty,x]} f_X(u)du).(\int_{]-\infty,y]} f_Y(v)dv) \\ &= F_X(x).F_Y(y) \end{aligned}$$

□

Remarque 5.1. *Dans la preuve ci-dessus la dérivée mixte $\frac{\partial^2}{\partial x \partial y}$ est légitime car toute fonction absolument continue a une dérivée presque partout.*

5.2 Lois de probabilité conditionnelle – Espérance mathématique conditionnelle

5.2.1 Lois de probabilité conditionnelle

Définition 5.2. *Soit (X,Y) un couple de variables aléatoires absolument continu. Soit f sa densité conjointe, f_x et f_Y ses densités marginales. Soit x un nombre réel tel que $f_X(x) \neq 0$.*

On appelle densité de Y conditionnelle à $\{X = x\}$, la fonction $f_{Y/X}(./x)$ de la variable y définie par $f_{Y/X}(y/x) = \dfrac{f(x,y)}{f_X(x)}$.

De même, soit un nombre réel y tel que $f_Y(y) \neq 0$, on appelle densité de X conditionnelle à $\{Y = y\}$, la fonction $f_{X/Y}(./y)$ de la variable x définie par : $f_{X/Y}(x/y) = \dfrac{f(x,y)}{f_Y(y)}$

Remarque 5.2. *Il résulte de la définition que pour presque tout x, $f(x,y) = f_{Y/X}(y/x).f_X(x)$ pour presque tout y.*

De même pour presque tout y, $f(x,y) = f_{X/Y}(x/y).f_Y(y)$, pour presque tout x. En effet si $f_X(x) > 0$, l'égalité est triviale. Supposons $f_X(x_0) = 0$, alors, on a $\int_{\mathbb{R}} f(x_0,y)dy = 0$, donc $y \longmapsto f(x_0,y) = 0$ pp, d'où, $f(x_0,y) = f_{Y/X}(y,x_0)f_X(x_0)$ pour presque tout y. On obtient la seconde relation en permutant les rôles de x et de y.

Remarque 5.3. *Il découle de la remarque précédente que :*

$$f_X(x) = \int_{\mathbb{R}} f(x,y)dy = \int_{\mathbb{R}} f_{X/Y}(x/y)f_Y(y)dy$$

$$f_Y(y) = \int_{\mathbb{R}} f(x,y)dx = \int_{\mathbb{R}} f_{Y/X}(y/x)f_X(x)dx$$

Proposition 5.4. *Pour tout x la fonction, $f_{Y/X}(./x) : y \longmapsto f_{Y/X}(y/x)$ est une densité de probabilité. De même pour tout y la fonction, $f_{X/Y}(./y) : x \longmapsto f_{X/Y}(x/y)$ est une densité de probabilité.*

Démonstration. $f_{Y/X}(y/x) \leq 0$ $\forall y$ et

$$\begin{aligned} f_X(x) &= \int_{\mathbb{R}} f(x,y)dy \\ &= \int_{\mathbb{R}} f(y/x)f_X(x)dy \\ &= f_X(x)\int_{\mathbb{R}} f_{Y/X}(y/x)dy \end{aligned}$$

Or $f_X(x) \neq 0\, pp \Longrightarrow \int_{\mathbb{R}} f_{Y/X}(y/x)dy$ pour presque tout x. Ainsi $f_{Y/X}(./x)$ est une densité de probabilité. La preuve est analogue pour $f_{X/Y}(./y)$. □

Proposition 5.5. *Soit (X,Y) un couple de variables aléatoires absolument continu. Soient f_X, f_Y, $f_{Y/X}(./x)$, $f_{X/Y}(./y)$, les densités marginales et conditionnelles. Si les variables X et Y sont indépendantes alors :*

1. *Pour tout x tel que $f_X(x) > 0$, on a : $f_{Y/X}(y/x) = f_Y(y)$.*
2. *Pour tout y tel que $f_Y(y) > 0$, on a : $f_{X/Y}(x/y) = f_X(x)$.*

Démonstration. Elle est triviale. □

5.2.2 Espérance mathématique conditionnelle

Définition 5.3. *Soit (X,Y) un couple de variables aléatoires absolument continu. On suppose que l'intégrale $\int_{\mathbb{R}} y f_{Y/X}(y/x)dy$ est absolument convergente. Alors,*

$$E(Y/X = x) = \int_{\mathbb{R}} y f_{Y/X}(y/x)dy$$

est appelée espérance mathématique de Y conditionnelle à $\{X = x\}$.

Remarque 5.4. *L'application $e : x \longmapsto E(Y/X = x)$ est une fonction réelle d'une variable réelle, mesurable et donc $e \circ X$ est une variable aléatoire réelle définie sur.*

Définition 5.4. *$e \circ X$ est appelée espérance mathématique conditionnelle de Y en X. On la note $E(Y/X)$.*

Théorème 5.1. *(de l'espérance mathématique conditionnelle)*
Soit (X/Y) un couple de variables aléatoires absolument continu. On suppose que $E(|Y|) < \infty$. Alors on a :

$$E(Y) = E[E(Y/X)]$$

Démonstration.

$$\begin{aligned} E(Y) &= \int_{\mathbb{R}} y f_Y(y)dy \\ &= \int_{\mathbb{R}} y\{\int_{\mathbb{R}} f(x,y)dx\}dy \\ &= \int_{\mathbb{R}} y\{\int_{\mathbb{R}} f_X(x).f_{Y/X}(y/x)dx\}dy \\ &= \int_{\mathbb{R}} \{\int_{\mathbb{R}} y f_{Y/X}(y/x)dy\}f_X(x)dx \\ &= \int_{\mathbb{R}} E(Y/X = x)f_X(x)dx \\ &= E[E(Y/X)] \end{aligned}$$

□

5.2.3 Courbe de régression

Définition 5.5. *Soit (X,Y) un couple de variables aléatoires vérifiant $E(|X|)<\infty$ et $E(|Y|)<\infty$. Alors le graphe de la fonction $x \longmapsto E(Y/X=x)$ est appelé la courbe de régression de Y en X. Le graphe de $y \longmapsto E(X/Y=y)$ est appelée courbe de régression de X et Y .*

Théorème 5.2. *Soit (X,Y) un couple de variables aléatoires absolument continu tel que $E(Y^2)<\infty$. La courbe de régression de Y en X vérifie la propriété de minimalité suivante : soit u une fonction réelle mesurable ; on suppose que l'expression $E[(Y-u(X))^2]$ est finie. Lorsque la fonction u varie, l'application $u :\longmapsto E[(Y-u(X))^2]$ atteint son minimum en $u(x)=E(Y/X=x)$. La valeur de ce minimum est $E[(Y-E(Y/X))^2]$.*

Démonstration.

$$\begin{aligned}
E[(Y-u(x))^2] &= \int_{\mathbb{R}^2}(y-u(x))^2 f(x,y)dxdy \\
&= \int_{\mathbb{R}} f_X(x)\{\int_{\mathbb{R}}(y-u(x))^2 f_{Y/X}(y/x)dy\}dx \\
&= \int_{\mathbb{R}} f_X(x)\{\int_{\mathbb{R}}(y^2-2yu(x)+u^2(x))f_{Y/X}(y/x)dy\}dx \\
&= \int_{\mathbb{R}} f_X(x)\{\int_{\mathbb{R}} y^2 f_{Y/X}(y/x)dy - 2yu(x)\int_{\mathbb{R}} y f_{Y/X}(y/x)dy + u^2(x)\int_{\mathbb{R}} f_{Y/X}(y/x)dy\}dx \\
&= \int_{\mathbb{R}} f_X(x)\{\int_{\mathbb{R}} y^2 f_{Y/X}(y/x)dy - 2yu(x)\int_{\mathbb{R}} y f_{Y/X}(y/x)dy + u^2(x)\}dx
\end{aligned}$$

$\int_{\mathbb{R}} f_{Y/X}(y/x)dy = 1$; ainsi :

$$E[(Y-u(X))^2] = \int_{\mathbb{R}} f_X(x)\{(u(x)-\int_{\mathbb{R}} y f_{Y/X}(y/x)dy)^2-(\int_{\mathbb{R}} y f_{Y/X}(y/x)dy)^2+\int_{\mathbb{R}} y^2 f_{Y/X}(y/x)dy\}$$

On voit que $E[(Y-u(X))^2]$ est minimale si :

$$u(x)=\int_{\mathbb{R}} y f_{Y/X}(y/x)dy = E(Y/X=x)$$

□

Remarque 5.5. *Les courbes de régression $x \longmapsto E(Y/X=x)$ et $y \longmapsto E(X/Y=y)$ sont en général distintes. Lorsque X et Y indépendantes, la courbe en x est une droite parallèle à Ox et celle en y est une droite parallèle à Oy.*

5.3 Règles de Calcul sur les espérances conditionnelles

Dans ce qui suit, nous écrirons indifféremment $E(Y/X)$ et $E_X(Y)$. Nous supposons aussi que toutes les espérances qui interviennent existent et que le couple (X,Y) est absolument continu.

©ENS/FAST/UP 2012

Définition 5.6. *Soit h une fonction réelle mesurable de deux variables réelles. On définit l'espérance mathématique de la variable aléatoire $h \circ (X, Y)$ conditionnelle à $\{X = x\}$ par :*

$$E(h \circ (X, Y)/X = x) = \int_{\mathbb{R}} h(x, y) f_{Y/X}(y/x) dy$$

Remarque 5.6. *Lorsque $h(x, y) = y$, on retrouve l'expression de $E(Y/X = x)$*

Théorème 5.3. *1. On a les identités suivantes :*

$$\begin{aligned} E[E(h \circ (X, Y))/X] &= \int E(h \circ (X, Y)/X = x) f_x(x) dx \\ &= E(h \circ (X, Y)) \end{aligned}$$

En prenant en particulier $h(x, y) = g(y)$, on obtient la formule de l'espérance conditionnelle pour la variable $g \circ Y$.

$$E[E(g \circ Y/X)] = \int E(g \circ Y/X = x) f_X(x) = E(g \circ Y)$$

2. Si X, Y sont indépendantes, alors , $E(g \circ Y/X) = E(g \circ Y)$.

3. On a toujours $E(g \circ X/X) = g \circ X$.

4. Pour X et Y quelconques, on a :

$$E_X(E_X(Y)) = E_X(Y)$$

$$E[(g_1 \circ X)(g_2 \circ Y)/X] = (g_1 \circ X) E(g_2 \circ Y/X)$$

Démonstration. 1. Posons $e(x) = E(h \circ (X, Y)/X = x)$.

$$E(h \circ (X, Y)/X) = e \circ X = e \circ \prod_1 \circ (X, Y)$$

$$\begin{aligned} E[E(h \circ (X, Y)/X)] &= E[e \circ X] \\ &= \int_{\mathbb{R}} e(x) f_X(x) dx \\ &= \int_{\mathbb{R}} E(h \circ (X, Y)/X = x) f_X(x) dx \\ &= \int_{\mathbb{R}} \{ \int_{\mathbb{R}} h(x, y) f_{Y/X}(y/x) dy \} f_X(x) dx \\ &= \int_{\mathbb{R}^2} h(x, y) f_{X,Y}(x, y) dx dy \\ &= E(h \circ (X, Y)) \end{aligned}$$

2. Si X et Y sont indépendantes, alors :

$$E(g \circ Y/X = x) = \int_{\mathbb{R}} g(y).f_{Y/X}(y/x) dy = \int_{\mathbb{R}} g(y).f_Y(y) dy = E(g \circ Y)$$

©ENS/FAST/UP 2012

$$e(x) = E(g \circ Y/X = x) = E(g \circ Y)$$

Implique que e est une fonction constante égale à $E(g \circ Y)$. Il en est alors de même de $e \circ X = E(g \circ Y/X)$ et donc :

$$E(g \circ Y/X) = E(g \circ Y)$$

.

3.

$$\begin{aligned} E(g \circ X/X = x) &= \int_{\mathbb{R}} g(x) f_{Y/X}(y/x) dy \\ &= g(x) \int_{\mathbb{R}} f_{Y/X}(y/x) dy \\ &= g(x) \end{aligned}$$

Ainsi en posant $e(x) = E(g \circ X/X = x)$, on a : $e = g$ et alors :

$$E(g \circ X/X) = e \circ X = g \circ X$$

4. En posant $g \circ X = E(Y/X) = E_X(Y)$, on a d'aprés 3),

$$E_X(E_X(Y)) = E_X(g \circ X) = E(g \circ X/X) = g \circ X = E_X(Y)$$

Soit

$$\begin{aligned} e(x) &= E[(g_1 \circ X)(g_2 \circ Y)/X = x] \\ &= \int_{\mathbb{R}} g_1(x) g_2(y) f_{Y/X}(y/x) dy \\ &= g_1(x) \int_{\mathbb{R}} g_2(y) f_{Y/X}(y/x) dy \\ &= g_1(x) E(g_2 \circ Y/X = x) \end{aligned}$$

$$E[(g_1 \circ X)(g_2 \circ Y)/X = x] = e \circ X = (g_1 \circ X) E(g_2 \circ Y/X)$$

□

Exemple 5.1. *Soit (X, Y) un couple de variables aléatoires absolument continues sur (Ω, τ, P). Evaluons $P(X < Y/X = x)$. Posons $A = \{X < Y\}$ et 1_A l'indicatrice de A.*

$$1_A = 1_{\{x<y\}} = h \circ (X, Y)$$

Avec $h(x, y) = 1_{\{x<y\}}(x, y)$. $P(A) = E(1_A)$.
En posant :

$$P\{A/X = x\} = E(1_A/X = x)$$

et

$$P\{A/X\} = E(1_A/X)$$

On a : $P(A) = \int_{\mathbb{R}} P(A/X = x) f_X(x) dx$.

Ainsi :

$$\begin{aligned} P\{X < Y/X = x\} &= E(1_{\{x<y\}/X=x}) \\ &= E(h \circ (X,Y)/X = x) \\ &= \int_{\mathbb{R}} h(x,y) f_{Y/X}(y/x)dy \\ &= \int_{\mathbb{R}} 1_{\{x<y\}}(x,y) f_{Y/X}(y/x)dy \\ &= \int_{\mathbb{R}} 1_{\{x<y\}}(y) f_{Y/X}(y/x)dy \\ &= \int_{\mathbb{R}} g(y) f_{Y/X}(y/x)dy \end{aligned}$$

En posant $g(y) = 1_{\{x<y\}}(y)$,

$$\begin{aligned} \int_{\mathbb{R}} g(y) f_{Y/X}(y/x)dy &= E(g \circ Y/X = x) \\ &= P(1_{\{x<y\}/X=x)} \\ &= P\{X < y/X = x\} \end{aligned}$$

Ainsi

$$P\{X < Y/X = x\} = P\{x < Y/X = x\}$$

Si les variables X *et* Y *sont indépendantes,* $f_{Y/X}(y/x) = f_y(Y)$, *et alors,*

$$\begin{aligned} P\{X < Y/X = x\} &= P\{x < Y/X = x\} \\ &= \int_{\mathbb{R}} 1_{\{x<y\}}(y) f_{Y/X}(y/x)dy \\ &= \int_{\mathbb{R}} 1_{\{x<y\}}(y) f_Y(y)dy \\ &= P\{x < Y\} \end{aligned}$$

Supposons X *et* Y *indépendantes et suivent la loi exponentielle de paramètre* λ *et* μ *respectivement.*

Les densités f_X *et* f_Y *de* X *et* Y *sont données par :*
$f_X(x) = \lambda e^{-\lambda x} 1_{\{x\geq 0\}}$ *et* $f_Y(y) = \mu e^{-\mu y} 1_{\{y\geq 0\}}$.

©ENS/FAST/UP 2012

$$\begin{aligned}
P\{X<Y\} &= \int_0^{+\infty} P\{X<Y/X=x\}f_X(x)dx \\
&= \int_0^{+\infty} P\{x<Y/X=x\}f_X(x)dx \\
&= \int_0^{+\infty} P\{x<Y\}f_X(x)dx \\
&= \int_0^{+\infty} e^{-\mu x}.\lambda e^{-\lambda x}dx \\
&= \lambda\int_0^{+\infty} e^{-(\mu+\lambda)x}dx \\
&= \frac{\lambda}{\lambda+\mu}
\end{aligned}$$

©ENS/FAST/UP 2012

Fonctions génératrices – fonctions caractéristiques

6.1 Fonction génératrice

Définition 6.1. *Soit X une variable aléatoire réelle telle que la fonction $g_X(u) = E(e^{uX})$ de la variable réelle u soit définie dans un voisnage ouvert de l'origine. Cette fonction est alors appelée fonction génératrice des moments de X.*

Remarque 6.1.

1. *g_X est toujours définie pour $u = 0$ et $g_X(0) = 1$.*
2. *Si X est à valeurs positives, alors g_X est continue et bornée sur $]-\infty, 0]$. Dans ce cas en faisant le changement de variable $u = -v$, on obtient la transformée de Laplace de X par :*
$$L_X(v) = g_X(-v) = E(e^{-vX})$$

Proposition 6.1. *Soit g_X la fonction génératrice des moments d'une variable aléatoire X.*

1. *Pour a, b réels, on a, $g_{aX+b}(u) = e^{bu} g_X(au)$.*
2. *La fonction $g_X(-u)$ est aussi une fonction génératrice des moments.*
3. *Si la loi de X est symétrique, c'est-à-dire si la loi de X est la même que la loi de $-X$, alors g_X est une fonction paire.*

Démonstration.

1. $g_{aX+b}(u) = E(e^{u(aX+b)}) = E(e^{ub}.e^{auX}) = e^{ub}E(e^{auX}) = e^{ub}g_X(au)$
2. $g_X(-u) = E(e^{-uX}) = E(e^{u(-X)}) = g_{-X}(u)$.
3. $g_X(u) = E(e^{uX}) = E(e^{u(-X)}) = E(e^{(-u)X}) = g_X(-u)$

□

Théorème 6.1. *(théorème d'unicité)*

Si deux variables aléatoires admettent une même fonction génératrice des moments, alors elles ont même loi.

Proposition 6.2. *Soit (X, y) un couple de variables aléatoires indépendantes dont chacune admet une fonction génératrice des moments. ALors, la variable aléatoire $X + Y$ admet une fonction génératrice desmoments et on a :*

$$g_{X+Y} = g_X.g_Y$$

©ENS/FAST/UP 2012

Démonstration.

$$g_{X+Y}(u) = E(e^{u(X+Y)}) = E(e^{uX}.e^{uY})$$

X et Y étant indépendantes e^{uX} et e^{uY} le sont aussi et alors,

$$E(e^{uX}.e^{uY}) = E(e^{uX}).E(e^{uY}) = g_X(u).g_Y(u)$$

□

Remarque 6.2. *La réciproque de la proposition précédente n'est pas vraie.*

Théorème 6.2. *Soit X une variable aléatoire admettant une fonction génératrice des moments g. Supposons que g soit définie pour tout u dans l'intervalle ouvert* $]u_1, u_2[$*, où* $u_1 < 0 < u_2$*. Alors,*

1. *Pour tout* $k \geq 1$*, où* $E(|X|^k) < \infty$*.*
2. *Pour tout* $t \in]-s, s[$ *avec* $0 < s < \min(-u_1, u_2)$*, on a :*

$$g(t) = 1 + E(X).\frac{t}{1!} + E(x^2)\frac{t^2}{2!} + ... + E(X^n)\frac{t^n}{n!} + ...$$

3. *Pour tout* $k \geq 1$*, on a :* $g^{(k)}(0) = E(X^k)$

Démonstration. Nous ferons la démonstration dans le seul cas où X admet une densité f.

1. $E(|X|^k) = \int_{\mathbb{R}} |x|^k f(x)dx$. Montrons que : $\int_{\mathbb{R}} |x|^k f(x)dx < \infty$. Soit s tel que $0 < s < u_0 = \min(-u_1, u_2)$. $e^{s|x|} = \sum_{k=0}^{+\infty} \frac{s^k|x|^k}{k!}$, et alors $\frac{s^k|x|^k}{k!} \leq e^{s|x|}\ \forall k \geq 1$. Ce qui implique $|x|^k \leq \frac{k!e^{s|x|}}{s^k},\ \forall k \geq 1$.

   Ainsi :

$$E(|x|^k) \leq \frac{k!}{s^k}\int_{-\infty}^{+\infty} e^{s|x|}f(x)dx = \frac{k!}{s^k}(\int_0^{+\infty} e^{sx}f(x)dx + \int_{-\infty}^{0} e^{-sx}f(x)dx)$$

$$E(|x|^k) \leq \frac{k!}{s^k}(\int_{-\infty}^{+\infty} e^{sx}f(x)dx + \int_{-\infty}^{+\infty} e^{-sx}f(x)dx) = \frac{k!}{s^k}[g_X(s) + g_X(-s)]$$

   $g_X(s)$ et $g_X(-s)$ étant finis alors, $E(|X|^k) < \infty$.

2.
$$g(t) = \int_{-\infty}^{+\infty} e^{tx}f(x)dx = \int_{-\infty}^{+\infty}(\sum_{k=0} +\infty\frac{(tx)^k}{k!})f(x)dx$$

   Posons $h(x) = f(x)\sum_{k=0}^{+\infty}\frac{(tx)^k}{k!} = e^{tx}.f(x)$ et $h_n(x) = f(x)\sum_{k=0} n\frac{(tx)^k}{k!}$ Il est évident que $\lim_{n\to+\infty} h_n(x) = h(x)$. Par ailleurs,

$$|h_n(x)| \leq f(x)\sum_{k=0} n\frac{|tx|^k}{k!} \leq f(x)\sum_{k=0} +\infty\frac{|tx|^k}{k!} = f(x)e^{|tx|} \leq f(x)e^{(tx)},\ \forall t \in]-s; s[$$

   De plus,

$$\int_{-\infty}^{+\infty} e^{s|x|}f(x)ds \leq g(s) + g(-s) < \infty$$

©ENS/FAST/UP 2012

On se trouve alors dans les conditions d'applications du théorème de la convergence dominée à $(h_n)_n$ et donc, on a :

$$\begin{aligned} g(t) &= \int_{-\infty}^{+\infty} h(x)dx \\ &= \lim_{n\to+\infty} \int_{-\infty}^{+\infty} h_n(x)dx \\ &= \lim_{n\to+\infty} \sum_{k=0} +\infty \frac{t^k}{k!} \int_{-\infty}^{+\infty} x^k f(x)dx \\ &= \lim_{n\to+\infty} \sum_{k=0} +\infty \frac{t^k}{k!} E(x^k) \ \forall t \in]-s,s[\end{aligned}$$

3. g étant développable en série entière dans $]-s,s[$, et on a : $g^k(0) = E(X^k)$.

□

Remarque 6.3. *D'après le théorème, si X admet une fonction génératrice, X admet des moments de tous les ordres. La réciproque est fausse comme l'illustre l'exemple suivant :*

Exemple 6.1. *Soit $X = e^Y$ ou Y suit une loi normale centrée, réduite $\mathcal{N}(0,1)$, c'est-à-dire la densité de Y est $g(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}}$, $\forall y \in \mathbb{R}$. La loi de X est dite log-normale (0,1).*

$$\begin{aligned} E(X^n) &= \int_{-\infty}^{+\infty} e^{ny} \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} dy \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{\frac{n^2}{2}} . e^{-\frac{(y-n)^2}{2}} dy \\ &= \frac{e^{\frac{n^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(y-n)^2}{2}} dy \\ &= e^{\frac{n^2}{2}} \end{aligned}$$

D'autre part,

$$g(u) = E(e^{uX}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{ue^y} . e^{-\frac{y^2}{2}} dy$$

Pour $u > 0$ et y assez grand $ue^y - \frac{y^2}{2} > 0$, d'où $e^{ue^y - \frac{y^2}{2}} > 1$ et alors

$$g(u) \leq \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} e^{ue^y - \frac{y^2}{2}} dy \leq \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} dy = +\infty$$

Donc pour $u \in]0,+\infty[$, $g(u)$ n'est pas définie et $]-\infty,0]$ n'est pas un voisinage ouvert de $u = 0$, d'où g n'est pas une fonction génératrice.

Définition 6.2. *Soit (X,Y) un vecteur aléatoire tel que la fonction $g(u,v) = E(e^{uX} + vY)$ des deux variables réelles u et v soit définie dans un voisinage ouvert de $(0,0)$. g est appelée fonction génératrice des moments de (X,Y).*

Remarque 6.4. *Le théorème d'unicité énoncé pour g_X est aussi valable pour g.*

Théorème 6.3. *Soit (X, Y) un vecteur aléatoire admettant une fonction génératrice des moments g. Alors,*

1. *Pour tous entiers $k \leq 1$, $l \leq 1$, on a : $E(|X|^k|Y|^k) < +\infty$.*
2. $$E(X^k.Y^k) = \frac{\partial^{k+l}}{\partial_u^k \partial_v^k} g(u,v)\Big|_{(u,v)} = (0,0)$$

Théorème 6.4. *Soit (X, Y) un vecteur aléatoire admettant une fonction génératrice des moments g. Alors, chacune des variables aléatoires marginales X, Y admet une fonction génératrice g_1, g_2. En outre, on a :*

$$g_1(u) = g(u, 0),\ g_2(u) = g(0, v)$$

Proposition 6.3. *Soit (X, Y) un vecteur aléatoire admettant une fonction génératrice des moments g. Alors, les deux propriétés suivantes sont équivalentes :*

1. *X et Y sont indépendantes.*
2. *$g(u, v) = g(u, 0).g(0, v)$.*

Démonstration. 1) $\Longrightarrow$ 2)
Supposons X et Y indépendantes, alors e^{uX} et e^{uY} sont indépendantes, et l'on a pour tout $(u, v) \in V$, voisinage ouvert de $(0, 0)$ sur lequel g est définie,

$$\begin{aligned} g(u,v) &= E(e^{uX+vY}) \\ &= E(e^{uX}.e^{vY}) \\ &= E(e^{uX}).E(e^{vY}) \\ &= g(u,0).g(0,v) \end{aligned}$$

2) $\Longrightarrow$ 1)
Supposons que 2) soit vérifié et soit $F(x, y)$ la fonction de répartition conjointe de (X, Y), $F_X(x)$, $F_Y(y)$ ses fonctions de répartition marginales.

$$g(u,v) = \int_{\mathbb{R}^2} e^{ux+vy} dF(x,y)$$

$$\begin{aligned} g(u,0).g(0,v) &= (\int_{\mathbb{R}} e^{ux} dF_X(x))(\int_{\mathbb{R}} e^{uy} dF_Y(y)) \\ &= \int_{\mathbb{R}^2} e^{ux+vy} dF_X(x) dF_Y(y) \end{aligned}$$

$$g(u, v) = g(u, 0).g(0, v),\ \forall (u, v) \in V \times V$$

En vertu du théorème d'unicité, on a : $F(x, y) = F_X(x).F_Y(y)$. Pour tout couple (x, y), d'où l'indépendance de X, Y.

□

Remarque 6.5. *Tout ce qui précède peut être généalisé à un vecteur aléatoire $(X_1, ..., X_D)$ donc la fonction génératrice est $g(u_1, u_2, ..., u_d) = E(e^{\sum_{i=1} du_i X_i})$.*

©ENS/FAST/UP 2012

6.2 Applications

6.2.1 La loi normale à deux dimensions

On dit qu'une variable aléatoire suit la loi normale de paramètre μ, σ et on note $\mathcal{N}(\mu, \sigma)$ si sa fonction densité est de la forme :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{1}{2}(\frac{x-\mu}{\sigma})^2, , \forall x \in \mathbb{R}$$

La loi normale $\mathcal{N}(0,1)$ est appelée loi réduite et admet pour densité :

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Soit (X'_1, X'_2) un vecteur aléatoire constitué de variables (X'_1), et (X'_2) indépendantes et de loi normale $\mathcal{N}(0,1)$. La fonction densité conjointe f est définie par :

$$f(x'_1, x'_2) = \frac{1}{2\pi} \exp(-\frac{1}{2}(x'^2_1 + x'^2_2))$$

Posons : $M' = \begin{pmatrix} X'_1 \\ X'_2 \end{pmatrix}$, $x' = \begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix}$,

$u' = \begin{pmatrix} u'_1 \\ u'_2 \end{pmatrix}$. La fonction génératrice de M' est : $g_{M'}(u') = E(e^{tu'M'})$, où $t_{u'} = (u'_1 u'_2)$.

$$\begin{aligned} g_{M'}(u') &= E(e^{tu'M'}) \\ &= \exp(-\frac{1}{2}(u'^2_1 + u'^2_2)) \\ &= \exp(\frac{1}{2}{}^t u'.u') \end{aligned}$$

De même,

$$f_{M'}(x') = \frac{1}{2\pi} \exp(-\frac{1}{2}(x'^2_1 + x'^2_2)) = \frac{1}{2\pi} \exp(\frac{1}{2}{}^t x'.x')$$

Si on associe à M' le couple $M = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ tel que $M = AM'$ où A est matrice carrée réelle d'ordre 2, alors on a :

1. $E(M) = E(AM') = AE(M') = A\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, donc M est centrée.
2. La matrice des variances-covariances $E(M - E(M))^t(M - E(M))$ de M est égale à $E(M^tM) = AE(M'^tM')^tA$. Or

$$M'^tM' = \begin{pmatrix} X'_1 \\ X'_2 \end{pmatrix}(X'_1, X'_2) = \begin{pmatrix} X'^2_1 & X'_1X'_2 \\ X'_1X'_2 & X'^2_2 \end{pmatrix}$$

Donc,

©ENS/FAST/UP 2012

$$\begin{aligned} E(M'^tM') &= \begin{pmatrix} E(X'^2_1) & E(X'_1X'_2) \\ E(X'_1X'_2) & E(X'^2_2) \end{pmatrix} \\ &= \begin{pmatrix} [E(X'_1)]^2 & E(X'_1).E(X'_2) \\ E(X'_1).E(X'_2) & [E(X'_2)]^2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= I \end{aligned}$$

Donc $E(M^tM) = AI^tA = A^tA$.

$$g_M(u) = E(e^{tuM}) = \exp(\frac{1}{2}{}^t u(A^tA)u),\ u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

Si en outre, A est une matrice régulière, M admet une densité conjointe,

$$f_M(x) = \frac{1}{2\pi}\exp(-\frac{1}{2}t_x(A^tA)^{-1}x)|detA^{-1}| \text{ où } x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Démonstration. $M = AM'$

$$\begin{aligned} g_M(u) &= E(e^{{}^tuM}) \\ &= E(e^{{}^tuA.M'}) \\ &= E(e^{{}^t({}^tAu)M'}) \end{aligned}$$

$$\begin{aligned} g_{M'}({}^tu{}^tAu) &= \exp[\frac{1}{2}{}^t({}^tAu).{}^tAu] \\ &= \exp[\frac{1}{2}{}^t u(A^tA)u] \end{aligned}$$

Si A est une matrice régulière, elle est inversible, et donc en faisant le changement de variable $\varphi : x \longmapsto x' = A^{-1}x = \varphi(x)$, alors on a :

$$f_M(x) = \frac{1}{2\pi}\exp(-\frac{1}{2}{}^t\varphi(x).\varphi(x))|\det\varphi|$$

où $\det\varphi$ appelée Jacobien de φ est égal à $\det A^{-1}$,

$${}^t\varphi(x) =^t (A^{-1}x) =^t x^t(A^{-1}) =^t x({}^tA)^{-1}$$

$${}^t\varphi(x)\varphi(x) =^t x({}^tA)^{-1}.A^{-1}x =^t x(A^tA)^{-1}x$$

d'où,

$$f_M(x) = \frac{1}{2\pi}\exp(-\frac{1}{2}{}^t x(A^tA)^{-1}x)|\det A^{-1}|$$

□

©ENS/FAST/UP 2012

Remarque 6.6. *La matrice A^tA est la matrice des variances - covariances du vecteur aléatoire $M = AM'$.*

Proposition 6.4. *Soit Γ une matrice des variances-covariances 2×2. Alors, il existe une matrice 2×2. Alors, il existe une matrice 2×2, réelle A telle que $\Gamma = A^tA$.*

Démonstration. Par définition Γ est réelle, symétrique, définie positive puisque ${}^tu\Gamma u = E(\|{}^tuM\|^2) \geq 0$. Elle admet donc des valeurs propres λ_1, λ_2 telles que $\lambda_1 \geq 0$, $\lambda_2 \geq 0$. Il existe une matrice 2×2 réelle orthogonal S telle que :

$$s^{-1}\Gamma S = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = \begin{pmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{pmatrix} . \begin{pmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{pmatrix}$$

Il en résulte que :
$\Gamma = S \begin{pmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{pmatrix} . \begin{pmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{pmatrix} S^{-1}$. D'où $\Gamma = A.{}^tA$, avec $A = S \begin{pmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{pmatrix}$.
Notons que S étant orthogonale, $S^tS = {}^t SS = I$, d'où $S^{-1} = {}^t S$

□

Définition 6.3. *Soit une matrice des variances-covariances* $\Gamma = \begin{pmatrix} \sqrt{\sigma_1}^2 & \rho\sqrt{\sigma_1}\sqrt{\sigma_2} \\ 0 & \sqrt{\lambda_2}^2 \end{pmatrix}$, $(\sigma_1 > 0, \sigma_2 > 0, |\rho| \leq 1)$. *On dit qu'un couple aléatoire,* $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ *suit la loi normale $\mathcal{N}_2(0, \Gamma)$, s'il admet pour fonction génératrice.*

$$g_M(u) = \exp(\frac{1}{2}{}^t u\Gamma u) = \exp(\frac{1}{2}(\sigma_1^2u_1^2 + 2\rho\sigma_1\sigma_2u_1u_2 + \sigma_2^2u_2^2))$$

Si en outre $\det\Gamma \neq 0$, *i.e si* $|\rho| < 1$, *il admet une densité conjointe donnée par :*

$$\begin{aligned} f_M(x) &= \frac{1}{2\pi}\exp(-\frac{1}{2}{}^t x\Gamma^{-1}x)x\sqrt{\det\Gamma^{-1}} \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp[-\frac{1}{2(1-\rho^2)}(\frac{x_1^2}{\sigma_1^2} - 2\rho\frac{x_1x_2}{\sigma_1\sigma_2} + \frac{x_2^2}{\sigma_2^2})] \end{aligned}$$

La loi $\mathcal{N}_2(0, \Gamma)$ est dite dégénérée ou non dégénérée selon que $\det\Gamma = 0$ *ou* $\det\Gamma \neq 0$. *Les deux propriétés suivantes découlent de la forme de $g_M(u)$.*

Propriété 6.1. *Les variables aléatoires marginales X_1, X_2 sont des variables aléatoires normales, centrées de variances respectives σ_1^2 et σ_2^2.*

Propriété 6.2. *Une condition nécessaire et suffisante pour que les variables aléatoires marginales X_1, X_2 soient indépendantes, est qu'elles soient non corrélées, c'est-à-dire $\rho = 0$. Ainsi dans le cas particulier des lois normales à deux dimensions , les propriétés d'indépendance et de non-corrélation de X_1, X_2 sont équivalentes.*

Théorème 6.5. *Soient* $M = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ *un couple aléatoire de loi $\mathcal{N}_2(0, \Gamma)$ et A une matrice 2×2 réelle. Alors, $M' = AM$ est un couple de loi $\mathcal{N}_2(0, \Gamma')$ où $\Gamma' = A\Gamma{}^tA$.*

 ©ENS/FAST/UP 2012

Démonstration.

$$\begin{aligned} g_{M'}(u') &= E(e^{tu'M'}) \\ &= E(e^{tu'AM}) \\ &= E(e^{{}^t(^tAu')M}) \\ &= g_M({}^tAu') \\ &= \exp(\frac{1}{2}{}^t({}^tAu')\Gamma{}^tAu') \\ &= \exp(\frac{1}{2}{}^tu'A\Gamma{}^tAu') \\ &= \exp(\frac{1}{2}{}^tu'(A\Gamma{}^tA)u') \end{aligned}$$

□

Remarque 6.7. *Si $\Gamma = I$, c'est-à-dire M est un couple de variables aléatoires indépendantes de loi $\mathcal{N}(0,1)$, alors le théorème rejoint le cas particulier de la proposition N° ?.*

Corollaire 10. *Si $M = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ est un couple indépendant de la loi $\mathcal{N}(0,1)$, alors le couple $M' = AM$ est indépendant de même loi, si et seulement si $A{}^tA = I$, c'est-à - dire que A est une matrice orthogonale.*

Remarque 6.8. *(Sur les lois dégénérées)*
La loi $\mathcal{N}_2(0,\Gamma)$, avec $\det\Gamma = 0$, c'est-à-dire $\rho = \pm 1$ n'admet pas de densité conjointe, mais admet une fonction génératrice obtenue en faisant $\rho = \pm 1$ dans l'expression générale de g_M. On a :

$$g_M(u) = \exp(\frac{1}{2}(\sigma_1 u_1 \pm \sigma_2 u_2)^2)$$

On voit que c'est la fonction génératrice d'un couple X_1, X_2 où $X_1 = \sigma_1 U$, $X_2 = \pm\sigma_2 U$, avec U une variable aléatoire de loi $\mathcal{N}(0,1)$.

Remarque 6.9. *(Sur les lois non centrées)*

Soit $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \in \mathbb{R}^2$ et M' un couple aléatoire de loi $\mathcal{N}_2(0,\Gamma)$. On note $\mathcal{N}_2(\mu,\Gamma)$ la loi de $M = M' + \mu$, et on l'appelle loi normale à deux dimensions de point moyen μ et de matrice de variances-covariances Γ. Sa fonction génératrice est donnée par :

$$g(u) = \exp({}^tu\mu + \frac{1}{2}{}^tu\Gamma u)\exp({}^tu\mu).g_{M'}(u)$$

Si en outre $|\rho| < 1$, elle admet une densité conjointe donnée par :

$$f_M(x) = \frac{\sqrt{\det\Gamma^{-1}}}{2\pi}\exp(-\frac{1}{2}{}^t(x-\mu)\Gamma^{-1}(x-\mu))$$

©ENS/FAST/UP 2012

6.3 Fonctions caractéristiques

Dans certains contextes, on préfère utiliser la fonction caractéristique d'une variable aléatoire plutôt que sa fonction génératrice des moments. La fonction caractéristique a l'énorme avantage d'exister sur $\mathbb{R}$ tout entier, et en outre, la correspondance entre loi de probabilité d'une variable aléatoire et sa fonction caractéristique est bijective.

Définition 6.4. *Soit X une variable aléatoire réelle. On appelle fonction caractéristique de X, la fonction de la variable réelle t définie par :* $\varphi(t) = E(e^{itX})$. *Si X est définie sur* (Ω, τ, P) *, on a :*

$$\varphi_X(t) = \int_\Omega e^{itX} dP = \int_{\mathbb{R}} e^{itx} dP_X$$

Cas particuliers

1. Si X est une variable aléatoire absolument continue de densité f, alors :

$$\varphi_X(t) = \int_{\mathbb{R}} e^{itx} f(x) dx$$

2. Si X est une variable aléatoire discrète prenant la valeur x_k avec la probabilité p_k, la loi de X est $\sum_k p_k S_k$ et alors $\varphi_X(t) = \sum_k p_k e^{itx_k}$.

Proposition 6.5. *Soit* φ_X *la fonction carastéristique d'une variable aléatoire X, alors :*

1. φ_X *est une fonction définie et continue pour tout nombre réel t.*
2. φ_X *est bornée, et en fait, pour tout t réel, on a :*

$$|\varphi_X(t)| \leq \varphi_X(0) = 1$$

3. $\varphi_X(-t) = \overline{\varphi_X(t)}$, $\forall t \in \mathbb{R}$, *en particulier si* φ_X *est réelle, alors* φ_X *est paire.*
4. *Pour tous a,b réels, on a pour tout réel t,*

$$\varphi_{aX+b}(t) = e^{ibt} \varphi_X(at)$$

5. *Si la loi de X est symétrique, alors* φ_X *est une fonction réelle, donc paire.*

Démonstration. 1. Pour montrer la continuité de φ_X en t, montrons que pour toute suite réelle $(t_n)_n$ tendant vers t, $\varphi_X(t_n)$ tend vers $\varphi_X(t)$.

$$\varphi_X(t_n) = \int_{\mathbb{R}} e^{it_n x} dP_X \text{ et } |e^{it_n x}| \leq 1$$

avec $\int_{\mathbb{R}} 1 dP_X = 1$. On se trouve dans les conditions d'application du théorème de la convergence dominé, et on obtient :

$$\lim_{n \to +\infty} \int_{\mathbb{R}} e^{it_n x} dP_X = \int_{\mathbb{R}} e^{itx} dP_X = \varphi_X(t)$$

Donc $\lim_{n \to +\infty} \varphi_X(t_n) = \varphi_X(t)$.

2. $$|\varphi_X(t)| = |\int_{\mathbb{R}} e^{itx} dP_X| \leq \int_{\mathbb{R}} |e^{itx} dP_X| = \int_{\mathbb{R}} dP_X = 1$$
   φ_X est donc bornée. En outre $\varphi_X(0) = 1$, d'où $|\varphi_X(t)| \leq \varphi_X(0)$, $\forall t \in \mathbb{R}$.
3. $\varphi_X(-t) = \overline{\varphi_X(t)}$ (évident).
   Si φ_X est réelle $\varphi_X(-t) = \varphi_X(t)$, d'où φ_X est paire.
4. Évident.
5. $\varphi_X(t) = \varphi_{-X}(t) = \varphi_X(-t)$, donc φ_X réelle, donc paire.

□

Théorème 6.6. *(Théorème d'unicité)*
Deux variables aléatoires admettant une même fonction caractéristique ont même loi.

Proposition 6.6. *Soit (X, Y) un couple de variables aléatoires indépendantes, alors :*

$$\varphi_{X+Y}(t) = \varphi_X(t).\varphi_Y(t)$$

Démonstration.

$$\varphi_{X+Y}(t) = E(e^{it(X+Y)}) = E(e^{itX}.e^{itY})$$

X et Y indépendantes, alors e^{itX} et e^{itY}, le sont donc,

$$E(e^{itX}.e^{itY}) = E(e^{itX}).E(e^{itY}) = \varphi_X(t).\varphi_Y(t)$$

□

Remarque 6.10. *La réciproque de la proposition précédente est fausse.*

Proposition 6.7. *1. Le produit de deux fonctions caractéristiques est une fonction caractéristique.*

2. Toute combinaison convexe de fonctions caractéristiques est une fonction caractéristique.

Démonstration. 1. Conséquence de la proposition N° ?.

2. Soient φ_X et φ_Y deux fonctions caractéristiques et p_1 et p_2 deux éléments de $[0, 1]$ tels que $p_1 + p_2 = 1$.

$$\begin{aligned} p_1\varphi_X(t) + p_2\varphi(t) &= p_1 \int_{\mathbb{R}} e^{itx} dP_X + p_2 \int_{\mathbb{R}} e^{ity} dP_Y \\ &= \int_{\mathbb{R}} e^{itx}(p_1 dP_X + p_2 dP_Y) \\ &= \int_{\mathbb{R}} e^{itx} d(p_1 P_X + p_2 P_Y) \end{aligned}$$

P_X et P_Y étant des probabilités $p_1P_X + p_2P_Y$ est une probabilité d'où le résultat. □

Remarque 6.11. *$\varphi_X + \varphi_Y$ n'est pas une fonction caractéristique, car $\varphi_X(0) + \varphi_Y(0) = 2$, au lieu d'être égal à 1. $\varphi_X - \varphi_Y$, n'est pas une fonction caractéristique, car, $\varphi_X(0) - \varphi_Y(0) = 0$, au lieu d'être égal à 1.*

Théorème 6.7. *Soit X une variable aléatoire de fonction caractéristique φ_X. Supposons qu'il existe un entier $n \geq 1$, tel que $E(|X|^n) < +\infty$. Alors :*

1. *φ_X est continûment dérivable jusqu'à l'ordre n inclus.*
2. *Pour tout $k = 0, 1, ..., n$, on a : $\varphi^{(k)}(0) = i^k E(X^k)$.*
3. *φ_X peut être représentée par la formule de Taylor :*

$$\varphi_X(t) = \sum_{k=0}^{n} \frac{(it)^k}{k!} E(X^k) + O(|t|^n)$$

Démonstration. 1. $\varphi_X(t) = \int_{\mathbb{R}} e^{itx} dP_X$ et $\dfrac{\partial^{(k)}}{\partial t} = i^k x^k e^{itx}$.

L'application : $t \longmapsto i^k x^k e^{itx}$ est continue pour tout k, et en outre, on a :

$$|i^k x^k e^{itx}| \leq |x|^k \text{ avec } \int_{\mathbb{R}} |x|^k dP_X = E(|x|^k) < +\infty, \ \forall k \in \{1, 2, ..., n\}$$

En conséquence, l'application $t \longmapsto (i)^k \int_{\mathbb{R}} x^k e^{itx} dP_X$ est continue $\forall k \in \{1, 2, ..., n\}$. Et en outre, on peut dériver k fois $\varphi_X(t)$, sous le signe somme et on a :

$$\varphi_X^{(k)}(t) = (i)^k \int_{\mathbb{R}} x^k e^{itx} dP_X$$

d'où φ_X est continûment dérivable jusqu'à l'ordre n inclus.

2. $\varphi_X^{(k)}(0) = (i)^k \int_{\mathbb{R}} x^k dP_X = i^k E(X^k)$
3. On applique à $\varphi_X(t)$ le développement de Taylor, et on a le résultat.

□

Fonction caractéristique d'un vecteur aléatoire

Soit $X = (X_1, ..., X_d)$ un vecteur aléatoire. On définit la fonction caraxtéristique φ_X de X par :

$$\varphi_X(t_1, ..., t_d) = E(e^{\sum_{i=1} dt_i X_i}) = \int_{\mathbb{R}^d} \prod_{i=1} de^{t_i x_i} dP_X$$

Proposition 6.8. *Soit (X, Y) un couple de variables aléatoires réelles dont la fonction caractéristique est notée $\varphi(t_1, t_2)$. Alors les deux propriétés suivantes sont équivalentes.*

1. *X et Y sont indépendantes.*
2. *Pour tout $(t_1, t_2) \in \mathbb{R}^2$, on a :*

$$\varphi(t_1, t_2) = \varphi(t_1, 0).\varphi(0, t_2)$$

Lois de probabilité de fonctions de variables aléatoires

7.1 Cas à une dimension

Théorème 7.1. *Soient I et I' deux intervalles ouverts, finis ou non de $\mathbb{R}$. Soit X une variable aléatoire réelle à valeurs dans I, absolument continue, de densité f; soient u bijection de classe $\mathcal{C}^1$ de I sur I' et u^{-1} sa bijection inverse. Alors $Y = u \circ X$ est une variable aléatoire à valeurs dans I', absolument continue de fonction densité g définie par :*

$$g(y) = f(u^{-1}(y))|(u^{-1})'(y)|1_{I'}(y)$$

Démonstration. Soit $A \in \mathcal{B}(\mathbb{R})$ avec $A \subset I'$.

$$P\{Y \in A\} = P\{X \in u^{-1}(A)\} = \int_{\mathbb{R}} f(x)1_{u^{-1}(A)}dx$$

En faisant le changement de variable $x = u^{-1}(y)$. On a :

$$P\{Y \in A\} = \int_{\mathbb{R}} f(u^{-1}(y))|(u^{-1})'(y)|1_A(y)dy$$

□

Exemple 7.1. *Soit X une variable aléatoire réelle de densité f. Alors, $Y = e^X$ est une variable aléatoire strictement positive dont la densité g est donnée par :*

$$g(y) = \frac{1}{y}f(Logy)1_{]0,+\infty[}(y)$$

En effet, $u : x \longmapsto e^x$ de $\mathbb{R}$ sur $\mathbb{R}_+^*$
$u^{-1} : y \longmapsto Logy$ de $\mathbb{R}_+^*$ sur $\mathbb{R}$, et donc
$u^{-1}(y) = \frac{1}{y}$.

Cas particulier

Supposons que X suit la loi $\mathcal{N}(\mu, \sigma)$, $(\mu \in \mathbb{R}, \sigma > 0)$, alors $Y = e^x$ (variable log-normale de paramètres μ, σ) a pour densité :

$$g(y) = \frac{1}{\sigma\sqrt{2\pi}}\frac{1}{y}\exp(-\frac{1}{2}(\frac{Logy - \mu}{\sigma})^2)1_{]0,+\infty[}(y)$$

©ENS/FAST/UP 2012

On rappelle que la fonction densité f de X est :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{1}{2}(\frac{x-\mu}{\sigma})^2)$$

Exemple 7.2. *On suppose que $Y = \frac{1}{X}$, où X est de densité f ,*

$$g(y) = \frac{1}{y^2} f(\frac{1}{y})$$

En effet
$$\begin{array}{cccc} u & : \mathbb{R}^* & \longrightarrow & \mathbb{R}^* \\ & x & \longmapsto & \frac{1}{x} \end{array} \qquad \begin{array}{cccc} u^{-1} & : \mathbb{R}^* & \longrightarrow & \mathbb{R}^* \\ & y & \longmapsto & \frac{1}{y} \end{array}$$

$$u^{-1}(y) = -\frac{1}{y^2} \Longrightarrow g(y) = f(u^{-1}(y))|(u^{-1})'(y)|1_{\mathbb{R}^*}(y) = \frac{1}{y} = f(\frac{1}{y}), y \neq 0$$

.

Cas particulier

Supposons que X suit la loi de Cauchy $C(0,1)$ et $Y = \frac{1}{X}$. $f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$, $\forall x \in \mathbb{R}$.

$$g(y) = \frac{1}{y^2} f(\frac{1}{y}) \frac{1}{y^2} \frac{1}{\pi} \frac{1}{1+\frac{1}{y^2}} = \frac{1}{\pi} \frac{1}{1+y^2}$$

.

On constate que Y suit encore la loi de Cauchy $C(0,1)$.

Remarque 7.1. *Dans le cas, où u n'est pas bijective, il n'y a pas de méthode générale pour déterminer la densité de $u \circ X$.*

Exemple 7.3. *Soit X une variable aléatoire réelle de densité f, et soit à déterminer la densité g de $Y = |X|$. La fonction $u(x) = |x|$ n'est pas bijective, donc la méthode précédente ne peut s'appliquer. Alors, on peut adopter la méthode suivante. Soit F_Y la fonction de répartition de Y.*

$$F_Y(y) = P\{Y < y\} = P\{|X| < y\} = \begin{cases} 0 & si\ y \leq 0 \\ P\{-y \leq X \leq y\} & si\ y > 0 \end{cases}$$

Ainsi

$$F_Y(y) = \begin{cases} 0 & si\ y \leq 0 \\ F_X(y) - F_X(-y) & si\ y > 0 \end{cases}$$

On en définit la densité f_Y de Y par dérivation,

$$f_Y(y) = \begin{cases} 0 & si\ y \leq 0 \\ f(y) + f(-y) & si\ y > 0 \end{cases}$$

Si la variable aléatoire X est paire, on a :

$$f_Y(y) = \begin{cases} 0 & si\ y \leq 0 \\ 2f(y) & si\ y > 0 \end{cases}$$

©ENS/FAST/UP 2012

Cas particulier

Si X suit la loi $\mathcal{N}(0,1)$, alors $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, et alors

$$f_Y(y) = \begin{cases} 0 & \text{si } y \leq 0 \\ \frac{2}{\sqrt{2\pi}} & \text{si } y > 0 \end{cases}$$

7.2 Cas à deux dimensions

Théorème 7.2. *Soient I et I' deux ouverts de $\mathbb{R}^2$, et (X,Y) un couple de variables aléatoires à valeurs dans I, de densité conjointe f. Soit :*

$$G(x,y) \longmapsto (u,v) = (u(x,y), v(x,y))$$

une bijection de classe $\mathcal{C}^1$ de I sur I' et soit

$$G^{-1} : (u,v) \longmapsto (x,y) = (h_1(u,v), h_2(u,v))$$

sa bijection inverse. Notons :

$$J = \frac{D(h_1,h_2)}{D(u,v)} = \frac{D(x,y)}{D(u,v)} = \begin{vmatrix} D_u h_1 & D_V h_1 \\ D_u h_2 & D_v h_2 \end{vmatrix}$$

Le Jacobien de bijection inverse G^{-1}. Alors,

$$(U,V) = G \circ (X,Y) = (u \circ (X,Y), v \circ (X,Y))$$

est un couple de variables aléatoires réelles à valeurs dans I' absolument continue de densité g définie par :

$$g(u,v) = f(h_1(u,v), h_2(u,v))|J| 1_{I'}(u,v)$$

Exemple 7.4. *Soit (X,Y) un couple de variables aléatoires indépendantes, dont chacune suit la loi $\mathcal{N}(O,1)$. Posons $U = \frac{X+Y}{\sqrt{2}}$ et $V = \frac{X-Y}{\sqrt{2}}$. Alors le couple (U,V) est formé de variables aléatoires indépendantes, dont chacune suit la loi $\mathcal{N}(O,1)$. En effet, on sait que la densité conjointe de (X,Y) est*

$$f(x,y) = \frac{1}{2\pi} \exp(-\frac{1}{2}(x^2+y^2))$$

Faisons le changement de variables $u = \frac{x+y}{\sqrt{2}}$ et $v = \frac{x-y}{\sqrt{2}}$.

L'application $(x,y) \longmapsto (u,v)$ est une bijection de $\mathbb{R}^2$ sur $\mathbb{R}^2$. La bijection inverse est définie par : $x = \frac{u+v}{\sqrt{2}}$ et $y = \frac{u-v}{\sqrt{2}}$. On a : $\frac{D(x,y)}{D(u,v)} = -1$, d'où la densité conjointe $g(u,v)$ de (U,V) est donnée par :

$$\begin{aligned} g(u,v) &= f(\frac{u+v}{\sqrt{2}}, \frac{u-v}{\sqrt{2}}).1 \\ &= \frac{1}{2\pi}[-\frac{1}{2}((\frac{u+v}{\sqrt{2}})^2 + (\frac{u-v}{\sqrt{2}})^2)] \\ &= (\frac{1}{2\pi} e^{-\frac{u^2}{2}}).(\frac{1}{2\pi} e^{-\frac{v^2}{2}}) \end{aligned}$$

©ENS/FAST/UP 2012

7.2.1 Loi de probabilité d'une fonction de deux variables aléatoires

Soient (X, Y) un couple de variables aléatoires réelles de densité conjointe f et $u : \mathbb{R}^2 \longrightarrow \mathbb{R}$ une fonction mesurable. On se propose de déterminer la loi de probabilité de la variable aléatoire $U = u \circ (X, Y)$. Pour ce faire, on suppose que l'application $(x, y) \longmapsto (u, v) = (u(x, y), y)$ est une bijection de classe $\mathcal{C}^1$ de $\mathbb{R}^2$ sur $\mathbb{R}^2$. Notons, $(u, v) \longmapsto (x, y) = (h(u, v), v)$ sa bijection unverse. Il en résulte que :

$$J = \frac{D(x,y)}{D(u,v)} = \begin{vmatrix} h'_u & h'_v \\ 0 & 1 \end{vmatrix}$$

D'aprés le théorème précédent (U, V) est absolument continu et sa densité conjointe g est définie par :

$$g(u, v) = f(h(u, v), v)|h'_u(u, v)| \ (u, v) \in \mathbb{R}^2$$

Par suite, on en déduit la loi de probabilité de U par sa densité :

$$\begin{aligned} g(u) &= \int_{\mathbb{R}} g(u, v)dv \\ &= \int_{\mathbb{R}} f(h(u, v), v)|h'_u(u, v)|dv \ u \in \mathbb{R} \end{aligned}$$

Exemple 7.5. *(Loi du produit)*
Supposons $U = XY$, alors $U = u \circ (X, Y)$, avec $u(x, y) = xy$. Posons $u = xy$, $v = y$, alors $x = \frac{u}{v}$, $y = v$.

$$J = \begin{vmatrix} \frac{1}{v} & -\frac{u}{v^2} \\ 0 & 1 \end{vmatrix} = \frac{1}{v}$$

et alors, $g(u) = \int_{\mathbb{R}} f(\frac{u}{v}, v)\frac{1}{|v|}dv \ \forall u \in \mathbb{R}$. Si le couple (X, Y) est indépendant, on a : $f(x, y) = f_1(x)f_2(y)$, et donc :

$$g(u) = \int_{\mathbb{R}} f(\frac{u}{v})f_2(v)\frac{1}{|v|}dv \ \forall u \in \mathbb{R}$$

Cas particulier

On suppose que X suit $\mathcal{N}(0, 1)$ et Y suit $\mathcal{N}(0, 1)$. Alors la densité g(u) de $U = XY$ est donnée par :

$$\begin{aligned} g(u) &= \frac{1}{2\pi}\int_{\mathbb{R}} \exp(-\frac{1}{2}(\frac{u^2}{v^2} + v^2))\frac{1}{|v|}dv \\ &= \frac{1}{\pi}\int_0^{+\infty} \exp(-\frac{1}{2}(\frac{u^2}{v^2} + v^2))\frac{1}{v}dv \end{aligned}$$

Exemple 7.6. *(loi du rapport)*
Supposons $U = \frac{X}{Y}$, $U = u \circ (X, Y)$ avec $u(x, y) = \frac{x}{y}$. Posons $u = \frac{x}{y}$, $v = y$, alors $x = uv$, $y = v$, et :

$$J = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = v$$

©ENS/FAST/UP 2012

La densité g de U est donnée par : $g(u) = \int_{\mathbb{R}} f(uv, u)|v|dv$. *Si* (X, Y) *est indépendant, on a :* $f(x, y) = f_1(x).f_2(y)$, *et*

$$g(u) = \int_{\mathbb{R}} f_1(uv) f_2(v)|v|dv$$

Cas particulier

X suit la loi $\mathcal{N}(0, 1)$ et Y suit $\mathcal{N}(0, 1)$. La densité g de $U = \frac{X}{X}$ est :

$$\begin{aligned} g(u) &= \frac{1}{2\pi} \int_{\mathbb{R}} \exp(-\frac{v^2}{2}(1 + u^2))|v|dv \\ &= \frac{1}{\pi} \int_0^{+\infty} \exp(-\frac{v^2}{2}(1 + u^2))vdv \end{aligned}$$

En faisant le changement de variable $\frac{1}{2}v^2(1 + u^2) = t$, on a : $vdv = \frac{dt}{1+u^2}$.

$$g(u) = \frac{1}{\pi} \frac{1}{1 + u^2} \int_0^{\infty} e^{-t} dt = \frac{1}{\pi} \frac{1}{1 + u^2} \; u \in \mathbb{R}$$

On constate que le rapport de deux variables aléatoires indépendantes normales, réduites, suit la loi de Cauchy $C(0, 1)$.

©ENS/FAST/UP 2012

Convergences stochastiques

8.1 Convergence en loi ou convergence étroite

Définition 8.1. *Soit $(X_n)_n$ une suite de variables aléatoires définies sur un espace probabilité (Ω, τ, P). Soit $(F_n)_n$, la suite de fonctions de répartition correspondantes. Soit X une variable aléatoire définie sur (Ω, τ, P), de fonction de répartition F. On dit que la suite $(X_n)_n$ converge en loi (ou étroitement) vers X si en tout point x où F est continue, $F_n(x)$ tend vers $F(x)$ lorque n tend vers $+\infty$. On note $X_n \xrightarrow{\mathcal{L}} X$ ou $\mathcal{L}(X_n) \longrightarrow \mathcal{L}(X)$.*

Autre terminologie :

On dit aussi que la suite des lois $\mathcal{L}(X_n)$ converge en loi ou étroitement vers la loi $\mathcal{L}(X)$.

Exemple 8.1. *Soit X_n uniformément répartie sur $[0, n]$ $(n \geq 1)$. La fonction de répartition correspondante F_n est donnée par :*

$$F_n(x) = \begin{cases} 0 & si\ x < 0 \\ \frac{x}{n} & si\ 0 \leq x < n \\ 1 & si\ x \geq n \end{cases}$$

$\lim_{n\to+\infty} F_n(x) = F(x) = 0$. *Or $F(x) = 0$ ne peut être la fonction de répartition d'aucune loi de probabilité, d'où $(X_n)_n$ ne converge pas en loi vers une limite.*

Exemple 8.2. *Soit $(X_n)_n$ une suite de variables aléatoires de loi $\mathcal{N}(\frac{1}{n}, \sqrt{1+\frac{1}{n}})$, $n \geq 1$, et soit X une variable aléatoire de loi $\mathcal{N}(0,1)$. Alors, la suite $(X_n)_{n\geq 1}$ converge en loi vers X. En effet :*

$$F_n(x) = \frac{1}{\sigma_n\sqrt{2\pi}} \int_{-\infty}^{x} \exp[-\frac{1}{2}(\frac{t-\mu_n}{\sigma_n})^2]dt \ \forall x \in \mathbb{R}$$

où $\sigma_n = \sqrt{1+\frac{1}{n}}$ et $\mu_n = \frac{1}{n}$. Posons $g = \frac{t-\mu_n}{\sigma_n}$, $dt = \sigma_n dz$, Pour $t \to -\infty$, $z \to -\infty$ et pour $t = x$, $z = \frac{x-\mu_n}{\sigma_n} = y_n$. Donc

$$\lim_{n\to+\infty} F_n(x) = \lim_{n\to+\infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty} y_n e^{-\frac{1}{2}z^2} dz$$

Or $y_n = \dfrac{x - \frac{1}{n}}{\sqrt{1+\frac{1}{n}}} \longrightarrow x$, quand $n \longrightarrow +\infty$. Donc :

©ENS/FAST/UP 2012

$$\begin{aligned}\lim_{n\to+\infty}\frac{1}{\sqrt{2\pi}}\int_{-\infty}y_ne^{-\frac{1}{2}z^2}dz &= \lim_{y_n\to x}\frac{1}{\sqrt{2\pi}}\int_{-\infty}y_ne^{-\frac{1}{2}z^2}dz\\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}xe^{-\frac{1}{2}z^2}dz\end{aligned}$$

D'où $(X_n)_n$ converge en loi vers X.

Exemple 8.3. *Soit la suite $(X_n)_{n\geq1}$ de variables aléatoires de loi $P_{X_n}=\frac{1}{2}(\delta_{-\frac{1}{n}}+\delta_{\frac{1}{n}})$, $n\geq1$. La suite $(X_n)_{n\geq1}$ converge en loi vers la variable aléatoire $X=0$. En effet,*

$$F_n(x)=P_{X_n}(]-\infty,x[)=\begin{cases}0 & si\ x<-\frac{1}{n}\\ \frac{1}{2} & si\ -\frac{1}{n}\leq x<\frac{1}{n}\\ 1 & si\ \frac{1}{n}\leq x\end{cases}$$

Il en résulte que pour tout réel x, on a :

$$F_n(x)\longrightarrow F^*(x)=\begin{cases}0 & si\ x<0\\ \frac{1}{2} & si\ x=0\\ 1 & si\ 0<x\end{cases}$$

F^* n'est pas une fonction de répartition, car elle n'est pas continue à droite de 0. Le point O est donc un point de discontinuité, et alors en prenant :

$$F(x)=\begin{cases}0 & si\ x<0\\ 1 & si\ x\geq0\end{cases}$$

On a $F^*=F$, sauf, en $x=0$, et donc pour tout $x\in C(F)$, (ensemble des points de discontinuité de F) on a :

$F_n(x)\longrightarrow F(x)$, et F est la fonction de répartition de $X=0$. Ce qui montre que (X_n) converge en loi.

Le théorème qui suit nous précise les liens entre convergence en loi d'un suite de variables alétoires et convergence de la suite des fonctions caractéristiques correspondantes.

Théorème 8.1. *(Paul Lévy)*

1. *Soit $(X_n)_n$ une suite de variables aléatoires qui converge en loi vers une variable aléatoire X. Soit $(\varphi_n)_n$ la suite des fonctions caractéristiques correspondantes ; alors $(\varphi_n)_n$ converge vers la fonction caractéristique φ de X et ceci uniformément dans tout intervalle fini.*
2. *Inversement, si la suite de fonctions caractéristiques d'une suite de variables aléatoires $(X_n)_n$ converge simplement vers une fonction φ de partie réelle continue à l'origine, aors :*
3. *il existe une loi de probabilité μ et une seule dont φ soit la fonction caractéritique.*
4. *La suite $(X_n)_n$ converge en loi vers μ.*

©ENS/FAST/UP 2012

8.2 Convergence en probabilité

Définition 8.2. *Soit $(X_n)_{n\geq 1}$, une suite de variables aléatoires définies sur un même espace probabilisé. (Ω, τ, P), et soit X une variable aléatoire définie sur le même espace. On dit que la suite $(X_n)_n$ converge en probabilité vers la variable aléatoire X, et l'on écrit $X_n \xrightarrow{P} X$ si :*

$$\forall \epsilon > 0, \lim_{n\to+\infty} P\{|X_n - X| > \epsilon\} = 0$$

Proposition 8.1. *(Unicité de la limite)*
Soient $(X_n)_n$ une suite de variables aléatoires et X, X' deux variables aléatoires telles que $X_n \xrightarrow{P} X$ et $X_n \xrightarrow{P} X'$. Alors, $P\{X = X'\} = 1$.

Démonstration. $\forall n$, on a : $|X - X'| \leq |X - X_n| + |X_n - X'|$. Soit N un entier positif, d'après l'inégalité ci-dessous, on a :

$$\{|X - X_n| < \frac{1}{2N}\} \cap \{|X_n - X'| < \frac{1}{2N}\} \subset \{|X - X'| \leq \frac{1}{N}\}$$

Ce qui implique,

$$P\{|X - X'| > \frac{1}{N}\} \leq P\{|X - X_n| \geq \frac{1}{2N}\} + P\{|X_n - X'| \geq \frac{1}{2N}\}$$

or

$$\lim_{n\to+\infty} P\{|X - X_n| \geq \frac{1}{2N}\} = \lim_{n\to+\infty} P\{|X_n - X'| \geq \frac{1}{2N}\} = 0$$

Donc $P\{|X - X'| > \frac{1}{N}\} = 0$.

Soit $A_N = \{|X - X'| > \frac{1}{N}\}, N \geq 1$. $(A_N)_N$ est une suite croissante et , $\bigcup_{N=1}^{+\infty} A_N = \{X \neq X'\}$.

$$P(\{X \neq X'\}) = P(\bigcup_{N=1}^{+\infty} A_N) = \lim_{n\to+\infty} P(A_n) = 0$$

d'où $P(\{X = X'\}) = 1$. □

Théorème 8.2. *Soit $(X_n)_n$ une suite de variables aléatoires telle que $X_n \xrightarrow{P} X$, $f : \mathbb{R} \longrightarrow \mathbb{R}$ une fonction continue, alors :*

$$f \circ X_n \xrightarrow{P} f \circ X$$

Démonstration. Il s'agit de montrer que pour chaque $\epsilon > 0$.

$$\lim_{n\to+\infty} P(\{|f(X_n) - f(X)| > \epsilon\}) = 0$$

Soit $B_N = \{|X| > N\}$, $N \geq 1$. La suite $(B_N)_{N\geq 1}$ est décroissante et
$\bigcap_{N=1} +\infty B_N = \phi$. Il en résulte que :

$$\lim_{n\to+\infty} P(B_N) = P(\phi) = 0$$

Ainsi, pour chaque $\delta > 0$, il existe un entier k_δ tel que $P(B_N) < \frac{\delta}{2}$ pour $N > k_\delta$. Fixons $N_0 > k_\delta$. La fonction f étant continue, elle est uniformément continue sur $[-N_0, N_0]$. Pour N_0

assez grand, soit ω_0 tel que $|X(\omega_0)| \leq N_0$. Alors, pour chaque $\epsilon > 0$, il existe un $\eta(\epsilon, N_0)$ et $X_n(\omega_0) \in [-N_0, N_0]$, alors, $|f(X_n(\omega_0)) - f(X(\omega_0))| < \epsilon$. Etant donné $\epsilon > 0$, posons :

$$A_n = \{|X| \leq N_0, |X_n| \leq N_0\},\ n \geq 1$$

$$A'_n = \{|X_n - X| \leq \eta(\epsilon, N_0)\},\ n \geq 1$$

$$C_n = \{|f(X_n) - f(X)| \leq \epsilon\},\ n \geq 1$$

D'après ce qui précède, $A_n \cap A'_n \subseteq C_n$ et alors,

$$C_n^c \subseteq A_n^c \cup A'^c_n\ n \geq 1$$

Il en résulte que :

$$P(C_n^c) \leq P(A_n^c) + P(A'^c_n),\ \forall n \geq 1$$

Mais $A_n^c \subset B_{N_0}\ \forall n \geq 1$, et alors,

$$P(C_n^c) \leq P(B_{N_0}) + P(A'^c_n),\ \forall n \geq 1$$

c'est-à-dire que

$$P(\{|f(X_n) - f(X)| > \epsilon\}) \leq P(B_{N_0}) + P(\{|X_n - X| > \eta(\epsilon, N_0)\})$$

Or $X_n \xrightarrow{P} X$ implique qu'il existe un entier k (dépendant de δ et $\eta(\epsilon, N_0)$) tel que pour $n \leq k$, on a :

$$P(\{|X_n - X| > \eta(\epsilon, N_0)\}) < \frac{\epsilon}{2}$$

En conséquence pour $n \geq k$, on obtient :

$$P(\{|f(X_n) - f(X)| > \epsilon\}) < \frac{\delta}{2} + \frac{\delta}{2} = \delta$$

puisque $P(B_{N_0}) < \frac{\delta}{2}$. □

Théorème 8.3. *Soit $(X_n)_n$ une suite de variables aléatoires. On suppose que $X_n \xrightarrow{P} X$ et que $P(\{X = 0\}) = 0$, alors : $\frac{1}{X_n} \overset{P}{\frac{1}{X}}$.*

Proposition 8.2. *Si la suite de points aléatoires $M_n = (X_n, Y_n)$, $n \geq 1$ converge en probabilité vers le point aléatoire $M = (X, Y)$ (c'est-à-dire, on a simultanément $X_n \xrightarrow{P} X$ et $Y_n \xrightarrow{P} Y$), alors on a :*

1. $X_n + Y_n \xrightarrow{P} X + Y$
2. $\lambda X_n \xrightarrow{P} \lambda X$, $(\lambda \in \mathbb{R})$
3. $X_n Y_n \xrightarrow{P} XY$
4. $\frac{X_n}{Y_n} \xrightarrow{P} \frac{X}{Y}$

Théorème 8.4. *(Critère de convergence en probabilité)*
Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires s'il existe $r > 0$ tel que la suite de terme général $E(|X_n|^r)$ tende vers O, alors la suite $(X_n)_n$ converge en probabilité vers O. La preuve utilisé le lemme suivant.

Lemme 8.1. *(Inégalité de Bienaymé-Tchébychev)*
Soit $r > 0$ un nombre réel et X une variable aléatoire réelle définie sur un espace probabilisé (Ω, τ, P). Si $E(|X|^r) < +\infty$, alors on a pour tout réel $t > 0$, l'inégalité :

$$P(\{|X| \geq t\}) \leq \frac{E(|X|^r)}{t^r}$$

Démonstration. (du lemme)

Soient $t, r > 0$, on a : $\{|X| \geq t\} \Longleftrightarrow \{|X|^r \geq t^r\}$. d'où

$$t^r.1_{|X|\geq t} = t^r.1_{|X|^r \geq t^r} \leq |X|^r$$

Il en résulte que :

$$E(t^r.1_{|X|\geq t}) = t^r E(1_{|X|^r \geq t^r}) \leq E(|X|^r)$$

C'est-à-dire que :

$$\int_\Omega t^r.1_{|X|\geq t} dP = t^r P\{|X| \geq t\} \leq E(|X|^r)$$

D'où $P\{|X| \geq t\} \leq \frac{E(|X|^r)}{t^r}$. □

Démonstration. (du théorème)
$P\{|X_n| \geq \epsilon\} \leq \frac{E(|X_n|^r)}{\epsilon^r} \longrightarrow 0$ quand $n \to +\infty$.

□

8.3 Convergence en moyenne d'ordre r

Définition 8.3. *Soit $(X_n)_n$ une suite de variables aléatoires définies sur un espace probabilisé (Ω, τ, P), et soit X une variable aléatoire définie sur le même espace. On dit que la suite $(X_n)_n$ converge vers X en moyenne d'ordre r si $E(|X_n - X|^r)$ converge vers 0, lorsque n tend vers $+\infty$ (Convergence dans L^r). Si $r = 2$, on parle de convergence en moyenne quadratique (Convergence dans L^2).*

Remarque 8.1. *$X_n \longrightarrow X$ en moyenne d'ordre r, cela n'implique pas que $E(|X|^r) < +\infty$. Mais, si $E(|X|^r) < +\infty$, alors $E(|X|^r) < +\infty \longrightarrow E(|X|^r) < +\infty$.*

8.4 Convergence presque sûre

Définition 8.4. *Soit $(X_n)_n$ une suite de variables aléatoires et X une variable aléatoire toutes définies sur (Ω, τ, P). On dit que la suite $(X_n)_n$ converge vers X presque surement, et on écrit $x_n \xrightarrow{PS} X$, s'il existe un ensemble P-négligeable A, tel que $X_n(\omega) \longrightarrow X(\omega)$, pour tout $\omega \in \Omega$ A. Dans certains cas, il est plus commode d'utiliser la définition équivalente suivante.*

Définition 8.5. *Soient $(X_n)_n$ et X des variables aléatoires définies sur (Ω, τ, P). Pour tout $\epsilon > 0$, posons,*

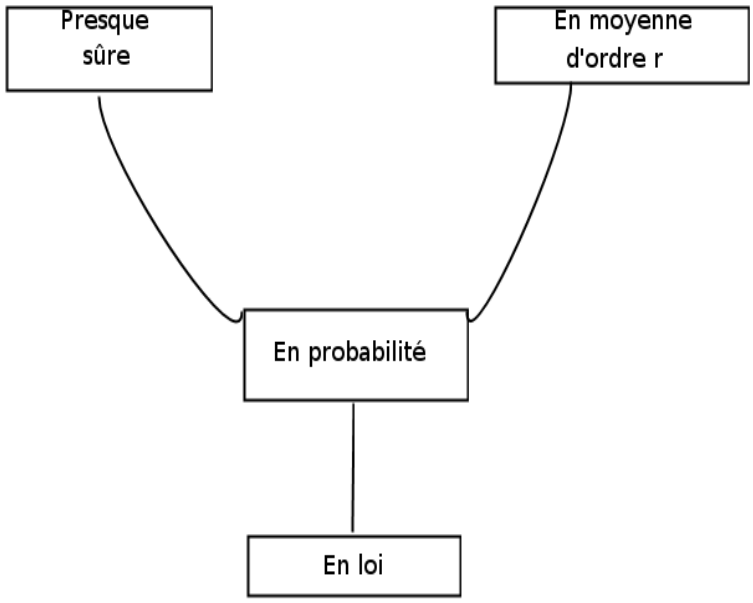
$$E_n(\epsilon) = \{|X_n - X| > \epsilon\},\ E(\epsilon) = \limsup_{n\to+\infty} E_n(\epsilon)$$

On dit que $(X_n)_n$ converge vers X presque surement si : $P(E(\epsilon)) = 0$.

Théorème 8.5. *Soit $(X_n)_n$ une suite de variables aléatoires. Les deux propriétés suivantes sont équivalentes :*

1. *La suite $X_n \xrightarrow{PS} 0$.*
2. *La suite de terme général $Y_n = \sup_{k\geq n} |X_k|$ converge vers 0 en probabilité.*

Remarque 8.2. *Il en résulte que la convergence presque sûre implique la convergence en probabilité.*



Exemple 8.4. *Posons*

$$E_k(\epsilon) = \{|X_k| > \epsilon\},\ E(\epsilon) = \limsup_{k\to+\infty} E_k(\epsilon)$$

$$\bigcup_{k\geq n} E_k(\epsilon) = \bigcup_{k\geq n} \{|X_k| > \epsilon\} = \{\sup_{k\geq n} |X_k| > \epsilon\}$$

$$E(\epsilon) = \inf_n \{\sup_{k\geq n} |X_k| > \epsilon\} = \lim_{k\to+\infty} \{\sup_{k\geq n} |X_k| > \epsilon\}$$

puisque la suite $\{\sup_{k\geq n} |X_k| > \epsilon\}_n$ est décroissante. En outre,

$$P(E(\epsilon)) = P(\lim_{k\to+\infty} \{\sup_{k\geq n} |X_k| > \epsilon\}) = \lim_{k\to+\infty} P\{\sup_{k\geq n} |X_k| > \epsilon\}$$

Ainsi,

$$P(E(\epsilon)) = 0 \Longleftrightarrow \lim_{k\to+\infty} P\{\sup_{k\geq n} |X_k| > \epsilon\} = 0$$

Théorème 8.6. *Soit $(X_n)_n$ une suite de variables aléatoiress telles que $X_n \xrightarrow{P} X$, alors il existe une suite partielle $(X_{nk})_{nk}$ extraite de $(X_n)_n$ telle que $X_{nk} \xrightarrow{PS} 0$.*

8.5 Comparaison des divers types de convergence

Théorème 8.7. *Soit $r > 0$, la convergence en moyenne d'ordre r entraîne la convergence en probabilité.*

Démonstration. D'après l'inégalité de Bienaymé-Tchebychev, on a, pour tout $\epsilon > 0$.

$$P(\{|X_n| \geq \epsilon\}) \leq \frac{E(|X_n|^r)}{\epsilon^r} \longrightarrow 0 \text{ lorsque } n \to +\infty$$

□

Théorème 8.8. *Si $(X_n)_n$ est une suite de variables aléatoires indépendantes, centrées de variances $VarX_n = \sigma_n^2 \forall n$, si la série $\sum_{n=1} \sigma_n^2$, est convergente alors, $\sum_{n=1} +\infty X_n$ est convergente presque surement. Le diagramme suivant schématise la comparaison entre les divers types de convergences.*

Exemple 8.5.

$$\begin{cases} X_n = -\frac{1}{n} \text{ avec la probabilité } \frac{n^2}{1+n^2} \\ X_n = n \text{ avec la probabilité } \frac{1}{1+n^2} \end{cases}$$

$$E(X_n) = \frac{n^2}{1+n^2} \times \frac{-1}{n} + \frac{n}{1+n^2} = 0$$

$$E(X_n^2) = \frac{n^2}{1+n^2} \times \frac{1}{n^2} + \frac{n^2}{1+n^2} = 1$$

Donc $(X_n)_n$ ne tend pas vers O en moyenne quadratique :

$$\{|X_n| > h\} = \{X_n < -h\} \cup \{X_n > h\}$$

Ainsi si $h > \frac{1}{n}$, alors, $-h < -\frac{1}{n}$ et $\{X_n < -h\} = \phi$. d'où $\{|X_n| > h\} = \{X_n > h\}$ et $P\{|X_n| > h\} = \frac{1}{1+n^2}$. Quand $n \to +\infty$, $P\{|X_n| > h\} \to 0$, donc X_n tend vers 0 en probabilité.

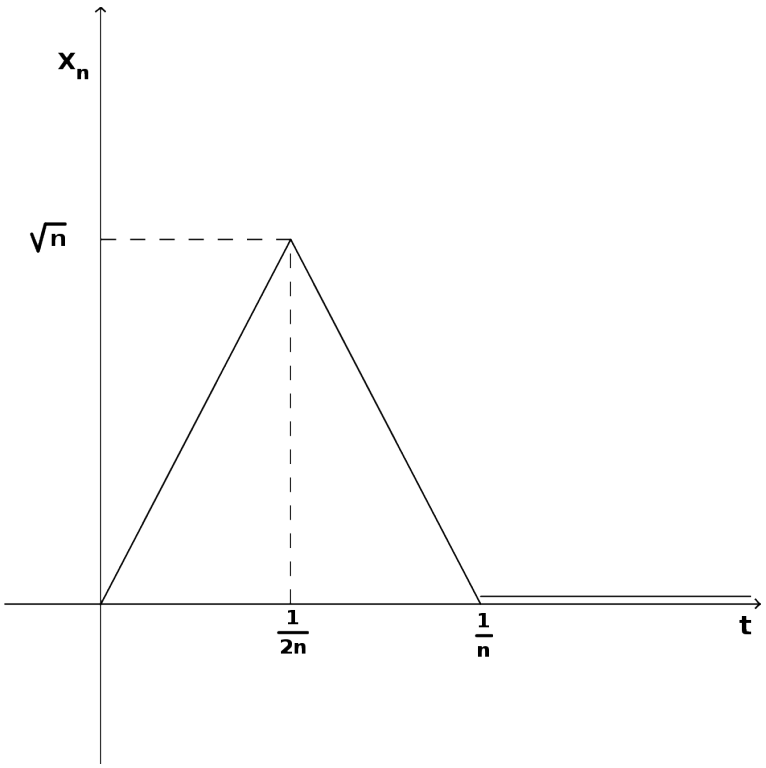
Exemple 8.6.

$$X_n(t) = \begin{cases} 2n^{\frac{3}{2}}t & si\ 0 \leq t < \frac{1}{2n} \\ 2\sqrt{n} - 2n^{\frac{3}{2}}t & si\ \frac{1}{2n} \leq t \leq \frac{1}{n} \\ 0 & si\ \frac{1}{n} \leq t \leq 1 \end{cases}$$

Lorque $n \to +\infty$, $X_n(t) \to 0$, pour tout t (c'est la convergence sûre).

$$E([X_n(t)]^2) = \int_0^1 ([X_n(t)]^2)dt = 2\int_0^{\frac{1}{2n}} (2n^{\frac{3}{2}}t)^2 dt = \frac{1}{3}$$

Donc, $E(X_n^2)$ ne tend pas vers O, d'où X_n ne tend pas vers O en moyenne quadratrique.



8.6 Loi des grands nombres

Définition 8.6. *La suite $(X_n)_{n\geq 1}$ satisfait la loi faible des grands nombres, si la suite de terme général $Y_n = \frac{1}{n}(X_1 + ... + X_n)$ converge vers 0 en probabilité. La suite $(X_n)_{n\geq 1}$ satisfait la loi forte des grands nombres, si la suite Y_n converge vers 0 presque surement.*

Théorème 8.9. *(Loi faible des grands nombres)*
Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires de L^2 (c'est-à-dire $E(|X^2|) < +\infty$) deux à deux non corrélés centrés. Pour tout n, on pose $VarX_n = \sigma_n^2$, $n \geq 1$. Si $\frac{1}{n^2}(\sigma_1^2 + ... + \sigma_n^2)$ tend vers 0, alors Y_n tend vers 0 en moyenne quadratique, et donc en probabilité.

Démonstration. $E(Y_n^2) = \frac{1}{n^2}(\sigma_1^2 + ... + \sigma_n^2)$ tend vers 0, donc Y_n tend vers 0 en moyenne quadratique, et par suite en probabilité. □

Théorème 8.10. *$(X_n)_{n\geq 1}$ une suite de variables aléatoires de L^2 deux à deux non corrélées. Pour tout $n \geq 1$, posons $E(X_n) = \mu_n$, et supposons que $\frac{1}{n}(\mu_1 + ... + \mu_2)$ tende vers μ, et que $\frac{1}{n^2}(\sigma_1^2 + ... + \sigma_n^2)$ tende vers 0. Alors, $Y_n = \frac{1}{n}(X_1 + ... + X_n)$ tend vers μ en moyenne quadratique et par suite en probabilité.*

Démonstration. La suite $(X_n - \mu)_{n\geq 1}$ est une suite de variables aléatoires centrées, donc d'après la loi faible des grands nombres :

$$\frac{1}{n}\sum_{k=1}^{n}(X_k - \mu_k) \longrightarrow 0$$

Or

$$\frac{1}{n}\sum_{k=1}^{n}(X_k - \mu_k) = \frac{1}{n}\sum_{k=1}^{n}X_k - \frac{1}{n}\sum_{k=1}^{n}\mu_k = Y_n - \frac{1}{n}\sum_{k=1}^{n}\mu_k$$

Donc $Y_n \longrightarrow \mu$ en moyenne quadratique.

□

Corollaire 11. *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires de L^2, deux à deux non corrélées, de même espérance mathématique μ, et de même variance σ^2, alors, $Y_n = \frac{1}{n}(X_1 + ... + X_n)$ tend vers μ, en moyenne quadratique, et donc en probabilité.*

Démonstration. On applique le théorème précédent avec : $\mu = \mu_1 = ... = \mu_n$.

□

Théorème 8.11. *(Loi forte des grands nombres)*
Soit $(X_n)_{n\geq 1}$, une suite de variables aléatoires de L^2, centrées et indépendantes. Pour tout $n \geq 1$, on pose $VarX_n = \sigma_n^2 < +\infty$. Si la série $\sum_{n=1}^{\infty} \frac{\sigma_1^2}{n^2}$ est convergente, alors la suite $(Y_n)_{n\geq 1}$ avec $Y_n = \frac{1}{n}(X_1 + ... + X_n)$ tend vers 0 presque surement.

Théorème 8.12. *(Rajchman)*
Soit $(X_n)_{n\geq 1}$, une suite de variables aléatoires de L^2, centrées et indépendantes. Pour tout $n \geq 1$, on pose $VarX_n = \sigma_n^2$. Si $\sup_n \sigma_n^2 < \infty$, alors :

1. *$Y_n \longrightarrow 0$ presque surement.*
2. *$Y_n \longrightarrow 0$ en moyenne quadratique.*

Démonstration. $\sigma^2 = \sup_n \sigma_n^2 < +\infty$, alors, $\sum_{n=1}^{+\infty} \frac{\sigma_n^2}{n^2} \leq \sum_{n=1}^{+\infty} \frac{1}{n^2}$ qui converge, d'où d'après la loi forte des grands nombres $Y_n \longrightarrow 0$ presque surement.

$$E(Y_n^2) = VarY_n = \frac{1}{n}\sum_{n=1}^{+\infty} \sigma_k^2 \leq \frac{\sigma^2}{n} \longrightarrow 0$$

d'où $E(Y_n^2) \longrightarrow 0$, c'est-à-dire $Y_n \longrightarrow 0$ dans L^2.

□

8.7 Théorème *"central limite"*

Il s'agit ici d'étudier dans quelles conditions une somme finie de variables aléatoires suit approximativement une loi normale.

Théorème 8.13. *(Lindeberg - Lévy)*
Soit $(X_n)_{n\geq 1}$, une suite de variables aléatoires undépendantes équidistribuées de L^2 (même fonction caractéristique, même loi). Posons :
$\mu = E(X_n)\, n \geq 1$, $\sigma^2 = VarX_n < +\infty$, $\forall n \geq 1$.
$S_n = X_1 + X_2 + ... + X_n$, $\bar{X_n} = \frac{S_n}{n}$,

$$Y_n = \frac{S_n n\mu}{\sigma\sqrt{n}} = \frac{\bar{X_n} - \mu}{\sigma/\sqrt{n}}$$

Alors, la suite $(Y_n)_{n\geq 1}$ converge en loi vers une variable aléatoire de loi $\mathcal{N}(0,1)$, c'est-à-dire pour tout réel x , on a :

$$\lim_{n\to+\infty} P(\{Y_n \leq x\}) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}\, du$$

Démonstration. Soit φ la fonction caractéristique de $X_1 - \mu$. En faisant le développement de Tayloy de φ, à l'ordre 2 au voisinage de 0, on a :

$$\varphi(t) = \varphi(0) + \frac{t}{1!}\varphi'(0) + \frac{t^2}{2!}\varphi''(0) + O(t^2) \text{avec } \lim_{t\to 0} O(t^2) = 0$$

Or

$$\varphi^{(k)}(0) = (i)^k E(X_1 - \mu)^k$$

Il en résulte que : $\varphi'(0) = iE(X_1 - \mu)$ et $\varphi'(0) = -\sigma^2$, et donc $\varphi(t) = \varphi(0) - \frac{t^2}{2}\sigma^2 + O(t^2)$. Or

$$Y_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \sum_{k=1}^{n} \frac{S_k - n\mu}{\sigma\sqrt{n}}$$

et donc

$$\varphi_{Y_n}(t) = (\varphi(\frac{t}{\sigma\sqrt{n}}))^n = (1 - \frac{t^2}{2n} + O(\frac{t^2}{n}))^n \longrightarrow e^{-\frac{t^2}{2}}$$

Ce qui établit le résultat en vertu du théorème de Paul Lévy.

□

Remarque 8.3. *L'importance du théorème réside dans le fait que pour n assez grand, les lois (en général compliquées) de $S_n, \bar{X}_n$ peuvent être approchées par les lois normales $\mathcal{N}(n\mu, \sigma\sqrt{n})$, et $\mathcal{N}(\mu, \sigma\sqrt{n})$.*

Exercice 8.1. *(Moivre-Laplace)*
Prenons pour loi commune des X_n, la loi de Bernoulli $pS_1 + pS_2(p, q > 0, p + q = 1)$. Alors S_n suit la loi binomiale $B(n, p)$. On a $E(S_n) = np$. $VarS_n = npq$, et la suite de variables aléatoires. $Y_n = \frac{(S_n - np)}{\sqrt{npq}}$ converge en loi vers une variable aléatoire de loi $\mathcal{N}(0, 1)$. Il en résulte que pour n assez grand $B(n, p)$ peut être approchée par la loi normale $\mathcal{N}(np, \sqrt{npq})$.

Exercice 8.2. *On suppose que les (X_n) suivent la loi de Poisson π_1 de paramètre 1, alors S_n suit la loi de Poisson π_n de paramètre n. On a : $E(S_n) = n$, $VarS_n = n$, et la suite de variables aléatoires $Y_n = \dfrac{S_n - n}{\sqrt{n}}$ converge en loi vers une variable aléatoire de loi $\mathcal{N}(0, 1)$.*

©ENS/FAST/UP 2012